

**Ein Beitrag zur
tonraumbasierten Analyse und Synthese
musikalischer Audiosignale**

Dissertation

zur Erlangung des akademischen Grades
Doktoringenieur (Dr.-Ing.)

vorgelegt der Fakultät für Elektrotechnik und
Informationstechnik der Technischen Universität Ilmenau

von Dipl.-Ing. Gabriel Gatzsche,
geboren am 21.05.1979 in Bad Langensalza

von Dipl.-Ing. Markus Mehnert,
geboren am 23.01.1976 in Hünfeld

Gutachter:

Prof. Dr.-Ing. Dr. rer. nat. h.c. Karlheinz Brandenburg

Prof. Dr.-Ing. Thomas Sporer

Prof. Thomas Krämer

vorgelegt am: 25.05.2009

Verteidigung am: 29.04.2011

Verf.-Nr.: EI 240

urn:nbn:de:gbv:ilm1-2011000164

Zusammenfassung

Das Ziel dieser Arbeit besteht darin, Verbesserungen in der Analyse und Synthese von Audiosignalen durch Anwendung von Tonräumen zu erreichen.

Im ersten Teil, der die Kapitel 2 bis 6 enthält und von Gabriel Gatzsche verfasst wurde, erfolgt die mathematisch-geometrische Beschreibung der Tonalität auf verschiedenen hierarchischen Ebenen angelehnt an Fred Lerdahls Tonal Pitch Space, David Gatzsches Kadenzkreis und Elaine Chew's Spiral Array (Berechnung von geometrischen Schwerpunkten in Tonraummodellen). Mit Hilfe zweier Formeln, der *Symmetriemodell-Generatorformel* und dem *SYM-Operator*, wird es möglich, die Entstehung der wichtigsten Hauptebenen der abendländischen Tonalität aus einer Quintreihe zu beschreiben, verschiedene, auf eine Tonart bezogene Modelle zu erzeugen und auf den jeweiligen Symmetrieton zu zentrieren. Damit gelingt es, eine Vielzahl bereits existierender Modelle zu verbinden und in ein einheitliches als *Symmetriemodell* bezeichnetes Framework zu integrieren. Um auch reale Musiksignale tonraumbasiert analysieren zu können, wird mit dem *Summenvektor im kreisförmigen Tonraum* ein Feature-Vektor vorgestellt, der wichtige tonale Eigenschaften eines Musiksignals niedrigdimensional repräsentiert. Dazu gehören z.B. funktionstheoretische Eigenschaften, das Tongeschlecht, Spannungs- und Auflösungsbestreben oder auch harmonische Mehrdeutigkeiten. Weiterhin wird der *Tonigkeits-Tonhöhenraum* eingeführt, der den unterschiedlichen Oktavlagen von Tonigkeiten geometrische Positionen so zuordnet, dass durch Wahl eines Raumausschnittes „gut klingende“ Akkorde erzeugt und durch Transformation des Raumausschnittes „günstig“ ineinander übergeblendet werden können. Dies führt zur Entwicklung eines neuartigen Musikinstrumentes, das als *HarmonyPad* bezeichnet wird. Dieses erlaubt einem Musiker, direkt mit geometrischen Tonräumen zu interagieren und damit Musiksignale zu erzeugen.

Markus Mehnert untersucht im zweiten Teil der Arbeit in den Kapiteln 7 bis 12 die Anwendbarkeit des Symmetriemodells auf konkrete Probleme des *Music Information Retrieval* (MIR). Hier werden sowohl die Tonart- als auch die Akkorderkennung betrachtet. Im Bereich der *Tonarterkennung*, die sich derzeit auf die Erkennung von Dur- und Molltonarten beschränkt, wird ein neuer Algorithmus vorgestellt, der auf dem Symmetriemodell basiert. Dieser verbessert den Stand der Technik erheblich. Darüber hinaus wird ein vollkommen neuer Ansatz vorgestellt, der die Tonarterkennung auf die sechs gebräuchlichsten Kirchentonarten erweitert, da diese besser als die reine Erkennung von Dur und Moll geeignet sind, den Charakter eines Musikstückes widerzuspiegeln. Zusätzlich wird ein neues Bewertungsmaß eingeführt, das den Vergleich mit zukünftigen Verfahren ermöglicht. Es wird ein für das MIR *neues maschinelles Lernverfahren* (HMM/KNN) vorgestellt, das die beiden Verfahren *Hidden Markov Models* und *k Nearest Neighbours* verknüpft. Im Bereich der *Akkorderkennung* werden mit diesem neuen Verfahren bessere Ergebnisse als mit allen vorhergehenden Verfahren erzielt. Dabei zeigt sich auch, dass der Merkmalsvektor des Symmetriemodells in Verbindung mit Akkorderkennung signifikant besser ist als der Chromavektor, der den Stand der Technik repräsentiert.

Abstract

The goal of the present work is to improve the analysis and synthesis of musical audio signals by the application of tonal pitch spaces.

The first part written by Gabriel Gatzsche consists of the Chapters 2 to 6. It discusses the mathematic-geometrical description of tonality on several hierarchical levels based on Fred Lerdahl's Tonal Pitch Space, David Gatzsche's Cadence Circle and Elaine Chew's Spiral Array (calculation of geometric centroids within tonal pitch spaces). Using two formulas, the *symmetry model generator formula* and the *SYM operator*, it is possible 1.) to describe the emergence of the most important levels of western tonality out of an array of fifths and 2.) to generate several key related models which are centered to the corresponding symmetry tone. With that steps it becomes possible to link several existing pitch spaces into a unified framework called *symmetry model*. To enable also the analysis of real music signals based on pitch spaces the *centroid vector within the circular pitch space* is introduced. This feature vector is a low dimensional representation of important tonal properties of musical audio signals. Such properties are functional relationships, the mode, tension and relaxation or harmonic ambiguities. Furthermore the *pitch class - pitch height space* is introduced. This space assigns geometric positions to different octaves of a given pitch class such that "well sounding" chords can be created by choosing a simple shaped region of the space. By transforming (rotating, translating, scaling etc.) such a region also well sounding chord transitions are generated. This leads to the development of a new musical instrument, called *HarmonyPad*. The HarmonyPad allows a musician to create music by interacting with pitch spaces directly.

Within the second part of the dissertation consisting of the Chapters 7 to 12 Markus Mehnert investigates the applicability of the symmetry model to concrete problems of *music information retrieval* (MIR) particularly chord and key recognition. The state of the art in the field of *key recognition* focuses on the estimation of major and minor keys. Within that work a new symmetry model based algorithm is presented which exceeds the results of current algorithms clearly. Additionally a new approach is proposed which extends key recognition to the estimation of the most often used six church modes. The latter represent the character of a musical piece in a better way than the standard modes "major" and "minor" do. Furthermore a new benchmark is introduced which allows the comparison of the current approach with future algorithms. A *new machine learning algorithm* (HMM/KNN) is proposed. The new algorithm combines the approaches *Hidden Markov Models* and *k Nearest Neighbours*. In the field of chord recognition the new approach achieves better results than all of the previous algorithms. It is shown that the symmetry model feature vector leads to significant better chord recognition results than the chroma vector which represents the state of the art.

Danksagung

Als erstes möchten wir Herrn *Prof. Karlheinz Brandenburg* als unserem Doktorvater und Betreuer danken. Trotz vielfältiger Aufgaben in der Fraunhofer-Institutsleitung und der Universitätsprofessur stand Prof. Brandenburg regelmäßig für Gespräche bereit und unterstützte uns mit Rat und Tat. Er gab uns die notwendigen Freiheiten und ein arbeitstechnisch exzellentes Umfeld, das zum Gelingen dieser Arbeit maßgeblich beitrug. Ein großer Dank gilt Herrn *Prof. Thomas Krämer*. Seine Harmonielehrebücher waren eine hervorragende Grundlage, uns tiefer in die klassische Harmonielehre einzuarbeiten und diese in das Symmetriemodell einzuordnen. Darüber hinaus gab er uns bei persönlichen Treffen wertvolle Hinweise und Anregungen, die wir in die vorliegende Arbeit einfließen lassen konnten. Ebenso möchten wir auch *Dr. Thomas Sporer* danken, der mit seiner Erfahrung und guten Ratschlägen auch dann weiter wusste, wenn der eine oder andere Hör- bzw. Usability-Test nicht ganz nach unseren Wünschen ausfiel. Seine positive Art lädt stets ein, Themen auch mal kontrovers zu diskutieren.

Vielen Dank an die Metadaten-Abteilung des Fraunhofer IDMT. Die dort entwickelte Transkription-Toolbox, welche die Grundlage für die Musikverarbeitung (Frontend) beinhaltet, stand uns uneingeschränkt zur Verfügung. Besonders bei Holger Großmann, Christian Dittmar, Matthias Gruhne und Daniel Gärtner möchten wir uns für die gute Zusammenarbeit bedanken.

Ein weiterer Dank gilt Herrn Prof. Dr. Eckard Lange von der Hochschule für Musik Franz Liszt, Weimar, dem wir unsere Ideen gerade in der so wichtigen Anfangszeit der Promotion vorstellen konnten. Durch ihn haben wir auch den Kontakt zu Prof. Krämer gefunden.

Gemeinsamer Dank gilt allen Studenten, die in Form von Diplomarbeiten, Medienprojekten und Hauptseminaren diese Arbeit unterstützten. Dazu gehören Tina Meyer, Jakob Abeßer, Thomas Mayenfels, Tobias Oeckler, Michael Büttner, Sebastian Schneider, Florian Raschke, Kristin Müller, Antje Storbeck, Arne Taute, Haojin Yang, Christian Stöcklmeier, Sebastian Bube, Jacob Korn, Fabian van de Sand, Daniel Arndt, Ting Zhao, Yawei Wang, Jan Bolz, Benjamin Hudarew, Felix Hirzel, Johannes Post, Michael Stein, Benjamin Schubert und Sascha Grollmich.

Persönlicher Dank Gabriel Gatzsche

Persönlich großer Dank gebührt meiner direkten Vorgesetzten Frau *Dr. Sandra Brix*, welche nicht nur mentale Unterstützung durch ihre offene und ermutigende Art bot. Vielmehr ermöglichte und förderte sie durch Entlastung von der täglichen Projektarbeit, durch Finanzierung von wissenschaftlichen Hilfskräften und Konferenzbesuchen und vielem mehr eine Arbeit, die anfänglich gar nicht in den Fokus der eigenen Abteilung zu passen schien. Gerade deshalb wünsche ich mir, die begonnenen Entwicklungen in erfolgreiche und marktfähige Produkte sowie die Erschließung neuer Forschungsfelder münden zu lassen. Besonderer Dank gilt auch Herrn *Dr. Thomas Sporer*, von dem ich vor allem lernte, gleichzeitig unterschiedliche Projekte im Blick zu behalten und auch Interesse für Themen zu zeigen, die über das Eigene hinausgehen. Eine besondere Seite der Doktorarbeit, an die ich immer wieder gerne zurückdenken werde, ist die Zusammenarbeit mit meinem Freund und Kollegen *Markus Mehnert*. Das intensive Diskutieren wissenschaftlicher Fragen, die gemeinsame Suche nach Lösungen, das Betreuen von Studenten, die Gründung des

Vereines AES Student Section Ilmenau e.V, das Zusammenschreiben der gemeinsamen Doktorarbeit, das Programmieren von Microcontrollern, die gemeinsame Mallorca-Arbeitsreise und vieles mehr haben mir unheimlich viel Spaß gemacht. Ohne die Zusammenarbeit mit Markus wäre eine Abdeckung des Analyse- als auch des Syntheseteils der vorliegenden Arbeit in dieser Breite niemals möglich gewesen.

Auch meinem Bruder *David Gatzsche* möchte ich danken, dessen Diplomarbeit einen der Ausgangspunkte der vorliegenden Arbeit bot. Danke für die vielen Gespräche, die Beantwortung musiktheoretischer Fragen, das Lesen des Manuskriptes sowie die Offenheit, das HarmonyPad musikpädagogisch einzusetzen.

Mein Dank gilt auch allen Kollegen der Abteilung Akustik, besonders meinen Freunden *Frank Melchior* und *Jens-Oliver Fischer*. Gemeinsames Mittagessen und die damit verbundenen Gespräche bildeten tägliche Höhepunkte. Durch Jens' privates Interesse für neuartige Musikcontroller und damit auch für das HarmonyPad ergaben sich immer wieder für mich inspirierende und motivierende Diskussionen.

Einen besonderen Dank möchte ich auch meiner Kollegin *Judith Liebetrau* aussprechen. Durch ihr unglaubliches Engagement im gemeinsamen EU-Projekt EDCine, durch die Vertretung des Teams auf den verschiedensten Projektmeetings und die selbstlose Übernahme vieler administrativer Aufgaben hielt sie mir an vielen Stellen den Rücken frei und trug so zu einem nicht unerheblichen Teil zum Gelingen der vorliegenden Arbeit bei.

In diesem Zusammenhang möchte ich auch Lutz Altmann danken, der im Rahmen seiner Diplomarbeit ebenfalls einen wichtigen Teil des EDCine-Projektes äußerst engagiert umsetzte und so dazu beitrug, dass auch dieses Projekt erfolgreich abgeschlossen werden konnte.

Dank auch allen anderen Kollegen der Abteilung Akustik, darunter Rene Rodigast, Katrin Reichelt, Carsten Land, Michael Strauß, Andreas Franck, Albert Zhykhar, Michael Beckinger, Clemens Clausens, Frank Fricke, Marko Doering, den Kollegen aller anderen Abteilungen sowie der freundlichen Verwaltung – Kathrin Höschel, Shirley Greulich & Team, Birgit Meder und Kerstin Remd, Kerstin Rose & Team, Sylvia Diettrich & Team), der IT-Abteilung (Jürgen Koller & Team).

Sehr genossen habe ich auch die Zusammenarbeit mit Dominik Strohmeier vom Institut für Medientechnik der TU Ilmenau. Seine Begeisterung für die Qualitätsbewertung von audiovisuellen Inhalten auf der Basis von Individual Profiling Ansätzen steckte uns sofort an und führte zu einem gemeinsamen Projekt, bei dem wir diese Ansätze auf die emotionale Charakterisierung von Akkordpaaren anwendeten. In diesem Zusammenhang möchte ich auch Andreas Ingerl vom Institut für Medienkommunikation danken, der diesem Projekt ebenfalls mit Rat und Tat zur Seite stand. Dank gilt auch Frau Prof. Krömker und Frau Julia Stumpf für die fachliche Beratung bei der Umsetzung der Usability Studie für das HarmonyPad.

Privat danken möchte ich meinen Eltern Eva und Andreas Gatzsche für ihre endlose Liebe, ihre Gebete, die stete Ermutigung, das Interesse und ihre Begeisterung für das Thema sowie die bedingungslose Annahme meiner Person als Mensch. Dank gilt Schwester Ruth Pohlmann, die ebenso wie meine Eltern oft betete, wenn ich nicht mehr weiter wusste.

Meinen WG-Kollegen Daniel Beer, Fabian Bethge, Jonathan Paulsen und Lothar Weigel danke ich für die tolle Zeit, die wir hier zusammen leben, Musik machen und beten konnten.

Herzlichen Dank auch an meine Großmutter Johanna Gatzsche, meine Tante Elisabeth Kohler, Frau Dr. Dagmar Mehnert-Vöcking sowie meinen Vater für das Korrekturlesen des Manuskriptes. Zutiefst dankbar bin ich auch für meine Verlobte Judith Abendroth. Ihre Liebe und Nähe sind für mich eines der größten Geschenke überhaupt. Die Perspektive, einmal mit ihr eine Familie zu gründen, haben meiner Arbeit unglaubliche Motivation gegeben.

Zuletzt, aber am allermeisten möchte ich meinem Gott und Vater im Himmel danken, der diese Erde mit all den wunderbaren Dingen geschaffen hat. Er hat uns Verstand gegeben, sie zu erforschen und zu durchdringen und der Menschheit nutzbar zumachen. Dank sei auch seinem Sohn Jesus Christus, in dessen Fußstapfen ich gehe und dem ich das unbeschreibliche Glück zu verdanken habe, schuldfrei, fröhlich und zukunftsgerichtet meinen Weg gehen zu können.

Persönlicher Dank Markus Mehnert

Ich möchte zuerst meinem verstorbenen Freund Dr. Horst Weißleder Danke sagen. Durch ihn habe ich nach dem Studium den Weg zurück an die TU Ilmenau gefunden mit dem Ziel, meine Dissertation zu verfassen. Ich habe viel von ihm gelernt und seine begeisterungsfähige Art wird mir in steter Erinnerung bleiben.

Bei meinen ehemaligen Kolleginnen und Kollegen Prof. Heidi Krömker, Prof. Hans-Peter Schade, Dr. Eckhardt Schön, Andreas Koch, Detlef Beyer, Bernd Hildenbrandt, Christian Weigel, Monique Zentgraf, Daniela Fiedler und Stephan Werner möchte ich mich für die angenehme und lehrreiche Zeit am Institut für Medientechnik herzlich bedanken.

Einen besonderen Dank möchte ich Daniel Arndt aussprechen. Er hat als einer meiner Diplomanden viel Zeit investiert, um meine Ideen, die ich fast täglich als Blockschaltbilder an eine große Tafel in den Kirchhoffbau Raum K3048 schrieb, in lauffähigen Quellcode umzusetzen. In diesem Raum entstand auch die Idee zum kombinierten HMM/KNN-Algorithmus, der in einer Vorversion bei der MIREX2008 den zweiten Platz - mit nicht signifikantem Abstand zum ersten - belegt hat.

Ich danke herzlich meinem Freund Gabriel Gatzsche. Wir haben uns in 2006 zufällig kennengelernt und recht schnell beschlossen, die Dissertation gemeinsam zu erarbeiten. In vielen, langen Diskussionen, die häufig bis in die Nacht hinein gingen, haben wir die Grundlagen des Symmetriemodells erarbeitet und immer wieder hinterfragt. Sein Bruder David Gatzsche war dabei eine große Hilfe. Auch bei ihm bedanke ich mich herzlich.

Ein weiterer Dank gilt Ania Lukashevich, Andreas Gatzsche, meiner Mutter Annemarie Mehnert, meinem Vater Walter Mehnert und meinem Schwager Michael Vöcking. Sie haben Teile des Manuskriptes gelesen und durch wertvolle Ratschläge zum Gelingen der Arbeit beigetragen.

Zuletzt möchte ich meiner Schwester Dr. Dagmar Mehnert-Vöcking besonders danken. Sie hat alle Versionen der Doktorarbeit Korrektur gelesen und auch so einige Wortdreher, die sich aus dem deutsch/englischen Berufsleben ergeben, wieder geradegebogen.

Vielen Dank Euch allen!

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Gliederung	4
2	Grundlagen der Tonalität	7
2.1	Tonalität	7
2.2	Das auditorische System des Menschen	9
2.3	Einzelöne	12
2.3.1	Das Obertonspektrum	12
2.3.2	Tonhöhe	12
2.3.3	Oktavähnlichkeit, Tonigkeit und Oktavlage	13
2.3.4	Quintähnlichkeit und Quintverwandtschaft	13
2.3.5	Identifikation bzw. Benennung von Tönen und Tonigkeiten	14
2.4	Intervalle	15
2.4.1	Identifikation bzw. Benennung von Intervallen	15
2.4.2	Intervallkonsonanz und -dissonanz	16
2.4.3	Grenzen der Intervallkonsonanz zur Erklärung der Harmonik	17
2.4.4	Satztechnische Bedeutung von Intervallen	18
2.5	Akkorde	19
2.5.1	Benennung von Dur- und Molldreiklängen	19
2.5.2	Akkordtypen	21
2.5.3	Wahrnehmung von Akkorden	21
2.6	Tonsets, Tonleitern, Modi und Skalen	23
2.6.1	Das diatonische Tonset	23
2.6.2	Das chromatische Tonset	25
2.6.3	Stimmung	25
2.6.4	Kirchentonleitern und weitere Skalen	27
2.6.5	Leitereigene und leiterfremde Klänge	28
2.7	Kontextbezogene Wahrnehmung von Tonalität	28
2.7.1	Tonale Hierarchien und tonale Stabilität	28
2.7.2	Wahrgenommene Nähe von Tönen und Akkorden	29

2.7.3	Wahrnehmung von Tonarten	31
2.7.4	Perzeptuelle Asymmetrien	33
2.7.5	Funktionstheoretische Aspekte	33
3	Stand der Wissenschaften Tonraummodelle	35
3.1	Heinichens, Matthesons und Kellners musikalische Kreise	36
3.2	Das Vial/Weber/Schönberg-Netz	37
3.3	Der Eulerraum	39
3.4	Das Euler/Riemann/Tonnetz	39
3.5	Der Terztorus	42
3.6	Elaine Chews Spiral Array	43
3.7	Fred Lerdahls Tonal Pitch Space	45
3.8	Das harmonische Heptagon/ Der Kadenzkreis	53
3.9	Generalized Voice-Leading Spaces	54
3.10	Dasian-, Guidonian- und Affinitätsräume	58
3.11	Weitere Arbeiten	61
3.12	Tonraumbasierte Analyse und Synthese von Musik	63
4	Das Symmetriemodell	67
4.1	Allgemeine Beschreibung kreisförmiger Tonräume	67
4.1.1	Verwendetes Koordinatensystem	68
4.1.2	Hilfstonigkeiten und Realtonigkeiten	68
4.1.3	Halbtonauflösung	68
4.1.4	Die Symmetrieachse und Symmetrietonigkeit	69
4.1.5	Grund- und erweiterter Index	69
4.1.6	Vektorrepräsentation von Tonigkeiten	70
4.1.7	Veränderte Indexzuordnung zu chromatischen Tonigkeiten	70
4.1.8	Zusammenfassung der Symbole	71
4.2	Die Teilräume des Symmetriemodells	72
4.2.1	Die Symmetriemodell-Generatorformel	72
4.2.2	Der tonartübergreifende Quintenzirkel F	73
4.2.3	Der tonartbezogene Quintenzirkel FR	73
4.2.4	Der tonartübergreifende Terzkreis T	75
4.2.5	Der tonartbezogene Terzkreis TR	75
4.2.6	Die diatonischen Tonleiterkreise D und DR	76
4.2.7	Die chromatischen Kreise C und CR	77
4.3	Erweiterungen des Symmetriemodells	78
4.3.1	Spiralförmige Darstellung des Symmetriemodells	78
4.3.2	Repräsentation chromatischer Tonigkeiten im FR , DR und TR	78
4.3.3	Alterationen	79
4.3.4	Repräsentation von Tonhöhe	80
4.4	Der Modulo- und der SYM-Operator	80

5	Musikalische Elemente im Symmetriemodell	87
5.1	Einführung	87
5.1.1	Summenvektoroperationen im kreisförmigen Tonraum	87
5.2	Grundelemente	89
5.2.1	Einzelne Tonigkeiten	89
5.2.2	Intervalle	91
5.2.3	Mehrklänge	92
5.2.4	Das diatonische Tonset	97
5.3	Funktionstheoretische Eigenschaften	98
5.3.1	Haupt- und Nebendreiklänge, Haupt- und Nebenfunktionen	98
5.3.2	Tonika-, Dominant- und Subdominantbereich	100
5.3.3	Charakteristische Zusatzdissonanzen	101
5.3.4	Verwandtschaftsgrade von Akkorden	102
5.3.5	Der Leitton, Strebetonigkeiten und Strebefähigkeit	103
5.3.6	Alterationen	104
5.4	Charakteristische Akkorde	106
5.4.1	Der verminderte Dreiklang	106
5.4.2	Der Dominantseptakkord	109
5.4.3	Zwischendominanten	112
5.4.4	Der Sixte ajoutée	112
5.4.5	Der verminderte Septakkord	113
5.4.6	Der übermäßige Dreiklang	115
5.5	Weitere Elemente	116
5.5.1	Kirchentonarten	116
5.5.2	Harmonisch Moll	116
5.5.3	Ausweichungen und Modulationen	117
5.5.4	Stabilität von Tönen in einem gegebenen Kontext	119
6	Tonraumbasierte Musiksynthese	123
6.1	Motivation	123
6.2	Aufbau eines Tonraumsynthesizers	125
6.3	Der Tonraum	125
6.3.1	Modellierung von Tonhöhe	127
6.4	Die Tonselektionsfunktion	128
6.5	Die Benutzerschnittstelle	131
6.5.1	Der Elo Touchscreen	131
6.5.2	Der JazzMutant Lemur	132
6.5.3	3D-SpaceNavigator	134
6.5.4	Nintendo Wiimote	136
6.6	Evaluierung	138
6.6.1	Testaufgaben	139

6.6.2	Verwendetes Parametermapping	140
6.6.3	Testaufbau	141
6.6.4	Testdurchführung	141
6.6.5	Ergebnisse	141
6.6.6	Diskussion	145
6.7	Anwendungsmöglichkeiten	147
7	Grundlagen Musikanalyse	149
7.1	Reales Audio – Frequenzanalyse	151
7.2	Frontend	153
7.2.1	Eigenschaften PCM-Signal	154
7.2.2	Auswahl des Zeitfensters	154
7.2.3	Constant-Q-Transformation	155
7.2.4	Tuning	157
7.2.5	Mapping	158
7.2.6	MIDI-Datenverarbeitung	158
7.2.7	Zusammenfassung der Parameter des Frontends	159
8	Grundlagen Maschinelles Lernen	161
8.1	Maschinelles Lernen allgemein	161
8.1.1	Überwachtes Lernen	162
8.1.2	Klassifizierung mittels Bayesscher Entscheidungstheorie	164
8.2	Verwendete Maschinelle Lernverfahren	165
8.2.1	KNN – k Nearest Neighbors	166
8.2.1.1	Klassifizierung bei KNN	167
8.2.2	HMM – Hidden Markov Models	169
8.2.2.1	Markov-Ketten	169
8.2.2.2	Modellparameter von HMM	170
8.2.2.3	HMM-Anwendungsfälle	172
8.2.2.4	Dekodierung – Klassifizierung bei HMM	173
8.2.3	HMM mit KNN-Schätzer	175
8.2.3.1	Klassifizierung mit KNN-Schätzer	176
9	Stand der Technik Musikanalyse	177
9.1	Tonarerkennung	177
9.1.1	Tonartprofile	178
9.1.2	Tonarerkennung mit symbolischem Audio	182
9.1.3	Tonarerkennung mit realem Audio	184
9.2	Akkorderkennung	189

10	Symmetriemodellparameter bei realem Audio	195
10.1	Repräsentation von Tönen, Intervallen und Akkorden	197
10.1.1	Einzelton	197
10.1.2	Intervall	198
10.1.3	Akkord	202
10.1.4	Grundkadenz	202
10.2	Repräsentation von Musikstücken	205
10.2.1	J. S. Bach „C-Dur Präludium“, BWV 846	205
10.2.2	L. van Beethoven „Für Elise“	205
10.2.3	W. A. Mozart „Eine kleine Nachtmusik“, KV 525	206
10.3	Zusammenfassung	210
11	Tonarerkennung	211
11.1	Erkennung von Dur- und Molltonarten	211
11.1.1	Algorithmus	212
11.1.1.1	Tonsetenergie	213
11.1.1.2	Winkelwichtung	214
11.1.1.3	Summenvektorbetrag in TR	216
11.1.1.4	Ermittlung des Tongeschlechts	216
11.1.2	Evaluierung	217
11.1.2.1	MIREX-Score	217
11.1.2.2	Parameter-Evaluierung	218
11.1.2.3	Ergebnis der Evaluierung	219
11.1.3	Vergleich mit anderen Verfahren	220
11.2	Erkennung von Kirchentonarten	225
11.2.1	Algorithmus	226
11.2.2	Evaluierung und Ergebnisse	228
11.2.2.1	Bewertungsmaßstab für die Erkennung von Kirchentonarten	228
11.2.2.2	Ergebnisse der Evaluierung	229
11.3	Zusammenfassung	230
12	Akkorderkennung	231
12.1	Gegenstand der Untersuchungen	231
12.1.1	Merkmalsvektoren und Analysealgorithmen	231
12.1.2	Ground Truth und verwendete Klassen	232
12.1.3	Training, Analyse und Berechnung der Ergebnisse	234
12.2	Akkorderkennung mit KNN	235
12.2.1	Evaluierung CHROMA	235
12.2.2	Evaluierung HSG6D	236
12.2.3	Evaluierung VSYM	237
12.2.4	Ergebnisse und Vergleich	238
12.3	Akkorderkennung mit HMM	240

12.3.1	Evaluierung VSYM	240
12.3.2	Ergebnisse und Vergleich	240
12.4	Akkorderkennung mit HMM/KNN	241
12.4.1	Evaluierung CHROMA	242
12.4.2	Evaluierung HSG6D	242
12.4.3	Evaluierung VSYM	243
12.4.4	Ergebnisse und Vergleich	244
12.5	Vergleich der Merkmalsvektoren und Algorithmen	245
12.5.1	Vergleich der Merkmalsvektoren	245
12.5.2	Vergleich der Analysealgorithmen	246
12.6	Vergleich mit dem Stand der Technik	247
12.7	Zusammenfassung	249
13	Zusammenfassung und Ausblick	251
13.1	Zusammenfassung	251
13.2	Ausblick	254
Verzeichnisse		1
Literaturverzeichnis		1
Abbildungsverzeichnis		21
Tabellenverzeichnis		25
Glossar und Abkürzungen		27
ANHANG		1
A	Längen der Summenvektoren von Intervallen	1
B	Beispiele zur Anwendung des SYM-Operators	7
B.1	Einordnung einer chromatischen Tonigkeit in die Quintreihe	7
B.2	Test, ob eine Tonigkeit Bestandteil eines diatonischen Tonsets ist	8
B.3	Einordnung einer Tonigkeit in einen der Unterräume des Symmetriemodells	8
B.4	Eigenschaften einer Tonigkeit in einem gegebenen diatonischen Tonset	9
B.5	Vorhersage des Auflösungsbestrebens einer Tonigkeit in einer gegebenen Tonart	9
B.6	Berechnung der Tonigkeitsbezeichnung einer Tonigkeit	9
B.7	Finden der einfachsten enharmonisch äquivalenten Tonart	10
C	Funktionssymbole	11
D	Mathematische Grundlagen	13
D.1	Statistische Auswertung von Ergebnislisten	13
D.2	Distanzmaße	15
E	Tonarerkennung: Verwendete Teststücke und deren Tonarten	17

F Ergebnisse Akkorderkennung

23

Kapitel 1

Einleitung

1.1 Motivation

Musik spielt in allen Kulturen der Welt eine große Rolle und ist einem permanenten Wandel unterzogen. Während die Musik der vergangenen Jahrhunderte hauptsächlich von einer inhaltlichen Weiterentwicklung geprägt war, wird die Musik der heutigen Zeit vor allem in der Art der Verbreitung, der Rezeption sowie der Art der Erzeugung einer drastischen Änderung unterzogen. Durch die *weltweite Vernetzung* ist zum einen der Zugriff auf nahezu unendliche Bestände von Musik möglich, zum anderen wird es immer einfacher, eigene Musikproduktionen einer großen Öffentlichkeit zur Verfügung zu stellen. Hier müssen Wege gefunden werden, gewünschte Musikstücke gezielt und schnell in vorhandenen Datenbeständen aufzufinden. Die *softwarebasierte Erzeugung und Bearbeitung von Musik* ermöglicht es darüber hinaus einem immer größer werdenden Personenkreis, selbst gestalterisch aktiv zu werden und musikalische Werke auf einem hohen Niveau zu produzieren. Dies führt erstens zu einem zunehmenden Bedarf an musikalischen Kreativwerkzeugen und zweitens spielt die Wiederverwendung existierender Inhalte anderer Menschen eine immer größere Rolle, was sich z.B. in der Entwicklung *neuartiger Lizenzmodelle* wie der Creative-Commons-Lizenz¹ widerspiegelt. Um jedoch die große Menge verfügbarer Inhalte auf Basis semantischer Merkmale sinnvoll rekombinieren zu können, müssen Verfahren entwickelt werden, welche in der Lage sind, multimediale Daten automatisiert inhaltsbasiert zu analysieren.

Ein anderer wichtiger Aspekt bezieht sich auf *elektronische Spiele*. Diese bieten realitätsnahe Grafik, professionellen Sound und fesseln den Spieler durch ausgeklügelte Belohnungs-

¹ <http://creativecommons.org/>, „Share, Remix, Reuse – Legally“

systeme. Verkaufszahlen von Spielen wie Guitar Hero (23 Millionen)² oder Sing Star (12 Millionen)³ zeigen, dass sich der heimische Spielcomputer mehr und mehr zu einem Musikinstrument weiterentwickelt. Während derzeitige Spiele oft nur die Kontrolle sehr weniger Parameter erlauben (wie z.B. das Aktivieren einer vorgegebenen Gitarrenstimme), wird schon bald die Nachfrage nach Spielen steigen, die auch die Manipulation und Gestaltung komplexerer musikalischer Eigenschaften erlauben. Hier müssen Methoden entwickelt werden, welche die wichtigsten Parameter von Musik der Wahrnehmung entsprechend auf einfach zu interpretierende Dimensionen abbilden und somit über einen Spielecontroller steuerbar machen.

Zuletzt sollen noch *Mobiltelefone* genannt werden. Diese haben sich schon lange als Audiospieler etabliert, werden immer kleiner, besitzen immer leistungsfähigere Prozessoren und verfügen über neuartige Eingabemethoden, wie z.B. Multi-Touch-Displays oder Beschleunigungssensoren. Ihre Verwendung als Musikinstrument wird dadurch immer interessanter. Hier sind jedoch Methoden erforderlich, musikalische Parameter auf engstem Raum (z.B. einem 3,5-Zoll-Display) zu organisieren und kontrollierbar zu machen. Eine Möglichkeit in diese Richtung ist der *Tonraum*. Der Tonraum ist eine in der Musiktheorie schon sehr lange verwendete Möglichkeit, Töne bzw. Tonfrequenzen geometrisch so zu organisieren, dass bestimmte musiktheoretische oder auch wahrnehmungspsychologische Eigenschaften klar aus der geometrischen Position der Töne hervorgehen. Ein sehr bekannter Tonraum ist der Quintenzirkel, welcher wichtige Verwandtschaftseigenschaften von Akkorden und Tonarten geometrisch repräsentiert. Der Begriff des Tonraumes wird schnell klar, wenn man ihn mit dem Farbraum, wie z.B. dem RGB-Raum, vergleicht. Dieser organisiert Farben entlang der Dimensionen rot, grün und blau. Wenn wir jedoch eine Menge von Personen bitten würden, Farben zu vergleichen, und die gewonnenen Ergebnisse anschließend mit Hilfe von Multidimensional Scaling analysieren würden, so käme heraus, dass Menschen Farben nicht entlang der Dimensionen R, G und B, sondern in den Dimensionen Sättigung, Helligkeit und Farbton wahrnehmen, was in einen anderen Farbraum, den HSV-Raum⁴, mündet. Ein solcher Raum ist deutlich intuitiver, denn er ermöglicht es, Farben in einer an der Wahrnehmung orientierten Weise zu vergleichen und damit auch zu manipulieren. Dasselbe gilt für Tonräume. Hier werden statt

² <http://www.computerbild.de/artikel/cbs-News-Bunt-gemischt-Aktuelle-Verkaufszahlen-Call-of-Duty-Guitar-Hero-3792752.html>, Stand 3.1.2009

³ http://www.cynamite.de/ps2/news/more/singstar_verkaufszahlen_aus_europa/9729/singstar_verkaufszahlen_aus_europa.html
Stand März 2008, Zahlen beziehen sich auf die PAL-Territorien)

⁴ HSV = Hue, Saturation, Value

Farben Töne räumlich so organisiert, dass sich entlang der jeweiligen Raumdimensionen bestimmte Wahrnehmungsparameter widerspiegeln. Je besser ein Tonraum organisiert ist, desto erfolgreicher können folgende Aufgaben gelöst werden:

- Vergleich von musikalischen Tönen, Akkorden, Musikstücken oder Sammlungen: Musikalische Strukturen, welche wahrnehmungsmäßig ähnlich sind, belegen benachbarte Regionen im Tonraum. Tonräume können damit helfen, musikalische Datenbestände zu organisieren, gezielter darauf zuzugreifen und einfacher zu durchsuchen.
- Vorhersage der zeitlichen und spektralen Gruppierung von Tönen: Damit ist die tonale Verschmelzung von Klängen ebenso gemeint wie die Formierung von Melodien und Akkorden. Diese Eigenschaft von Tonräumen kann für die Entwicklung neuartiger Musikinstrumente oder elektronischer Spiele ausgenutzt werden. Töne, die gut zueinander passen, liegen im Tonraum benachbart. Töne, die nicht miteinander gespielt werden sollen, liegen entfernt. Dadurch können potentielle Bedienfehler beim Spiel eines entsprechenden Instrumentes reduziert werden, und es wird möglich, kleinere Instrumente zu entwickeln. Dies ist Voraussetzung für die Weiterentwicklung mobiler Endgeräte, wie z.B. Mobiltelefone oder Kleincomputer, zu Musikinstrumenten.
- Bereitstellen einer Verbindung zwischen musikalischer Struktur und musikpsychologischer Wahrnehmung: Dazu können Emotionen, Assoziationen und auch visuelle Analogien gehören. Auf der Basis von Tonräumen ist es möglich, die musikpsychologische Wirkung von Tonkombinationen vorherzusagen und auf der Grundlage dieser Informationen wiederum Musikstücke zu vergleichen oder auch Werkzeuge anzubieten, um Musik eines bestimmten Charakters zu erzeugen.

Die Ziele dieser Arbeit bestehen deshalb darin,

1. geeignete Tonraummodelle auszuwählen bzw. zu entwerfen,
2. Werkzeuge zu schaffen, welche es ermöglichen, Musiksignale auf Basis von Tonräumen zu erzeugen und
3. Verfahren zu entwickeln, welche Musiksignale mit Hilfe von Tonräumen analysieren.

Während sich der letzte Punkt (Analyse) in der Entwicklung einer leistungsfähigen Tonart- und Akkorderkennung manifestieren wird, hat der zweite Punkt (Synthese) die

Entwicklung eines neuartigen auf Tonräumen basierten Instrumentenkonzeptes zum Ziel. Die hier entwickelten Techniken können in vielen Bereichen Anwendung finden, darunter dem Vergleich und der Suche von Musikstücken, der Identifikation von Cover-Songs, der automatisierten Transkription von Musikstücken, der Entwicklung von inhaltsmerkmalbasierten Kreativwerkzeugen oder auch der Entwicklung neuartiger elektronischer Musikspiele, darunter Logik-, Reaktions- und Geschicklichkeitsspiele.

Insgesamt erhoffen sich die Autoren durch die vorliegende Arbeit nicht nur die Grundlage für bessere Musikvergleichs- bzw. -suchverfahren zu legen, sondern auch Voraussetzungen zu schaffen, um aktives Musizieren einer deutlich größeren Anzahl von Menschen zu ermöglichen. Während es den meisten Menschen leicht fällt, Musik in wichtigen Grundaussagen zu verstehen, die darin verborgenen Emotionen zu erfassen, Spannung und Entspannung, Konsonanz und Dissonanz zu erleben, ist es nur sehr wenigen Menschen gegeben, eigene musikalische Ideen in die Realität umzusetzen. Gerade dieses eigene Musizieren fördert die Entwicklung sowohl der so wichtigen kommunikativen als auch geistig-logischen Fähigkeiten [74], [141]. Der Tonraum und die damit verbundenen Entwicklungen sollen ein wichtiger Schritt zur Förderung von musikalischer Kreativität sein.

1.2 Gliederung

- In Kapitel 2 werden zunächst die zum Verständnis von geometrischen Tonalitätsmodellen erforderlichen Grundlagen eingeführt. Neben musikalisch-musiktheoretischen und psychoakustischen Grundlagen von Tönen, Intervallen, Akkorden und Tonleitern spielen auch wahrnehmungspsychologische Aspekte der Tonalität eine Rolle.
- Kapitel 3 stellt existierende Tonraummodelle vor, angefangen bei Modellen der Generalbasszeit des 17. und 18. Jahrhunderts bis hin zu Modellen der Gegenwart. Neben der Darstellung der Modelleigenschaften spielt die Bewertung der Modelle hinsichtlich einer möglichen Einsetzbarkeit in der Audioanalyse und -synthese eine wichtige Rolle.
- In Kapitel 4 wird das Symmetriemodell als ein neuartiges geometrisches Tonalitätsmodell eingeführt. Das Symmetriemodell integriert eine Reihe von unabhängig existierenden Tonraummodellen in ein einheitliches Framework und zeigt einen Weg, den Zusammenhang unterschiedlicher hierarchischer Ebenen von Tonalität mathematisch sehr effektiv zu beschreiben. Mit der Einführung des Summenvek-

tors im kreisförmigen Tonraum wird darüber hinaus eine Möglichkeit aufgezeigt, kreisförmige Tonräume zur Analyse von Audiosignalen einzusetzen.

- Das Ziel von Kapitel 5 besteht darin, wichtige musiktheoretische Grundelemente in das Symmetriemodell einzuordnen. Neben der Repräsentation funktionstheoretischer Eigenschaften von Tönen und Akkorden werden auch erweiterte Aspekte wie Alteration, tonale Stabilität oder Leit- und Strebefähigkeit von Tönen behandelt.
- Viele musiktheoretische Eigenschaften von Tönen lassen sich in Tonräumen und besonders im Symmetriemodell sehr einfach und klar repräsentieren. Damit ist es auch möglich, bestimmte musikalische Ton- und Akkordkombinationen durch Anwendung von Tonräumen zu erzeugen. In Kapitel 6 werden deshalb Möglichkeiten zur Auralisierung von Tonräumen aufgezeigt. Mit dem HarmonyPad wird darüber hinaus ein neuartiges tonraumbasiertes Musikinstrument vorgestellt, welches auf dem Symmetriemodell basiert und die in den vorherigen Kapiteln entwickelten Grundlagen praktisch anwendet.
- In Kapitel 7 werden die Grundlagen der Musikanalyse erläutert. Das Musiksignal wird im Frontend in eine geeignete Repräsentation (Reihe von Merkmalsvektoren) umgewandelt, im ersten Schritt in den Chromavektor. Dieser wird entweder direkt verwendet oder in ein anderes Tonraummodell überführt. Ein Erkennungsalgorithmus nutzt anschließend die Merkmalsvektoren, um die eigentliche Musikanalyse durchzuführen.
- Das Kapitel 8 gibt zunächst einen allgemeinen Überblick über maschinelles Lernen (Analysealgorithmen). Anschließend werden die in dieser Arbeit verwendeten maschinellen Lernverfahren gegenübergestellt und diskutiert. Dabei wird ein für die Musikanalyse neues Verfahren eingeführt, das *Hidden Markov Models* und *K Nearest Neighbours* in einem Verfahren verbindet.
- Um die hier neu vorgestellten Ansätze vergleichen und einordnen zu können, wird in Kapitel 9 ein umfassender Überblick zum Stand der Technik in der Tonarterkennung und Akkorderkennung gegeben. Die Tonarterkennung unterteilt sich in die Analyse von Partituren (symbolisches Audio) und echten Audiosignalen (reales Audio). Bei der Analyse von realem Audio werden Tonartprofile von Krumhansl und Temperley eingesetzt. Zur Akkorderkennung werden hauptsächlich *Hidden Markov Models* verwendet.
- In Kapitel 10 werden die theoretischen Überlegungen zum Symmetriemodell mit echten Audiosignalen belegt. Es werden aufgezeichnete Töne, Intervalle und Akkor-

de im Symmetriemodell geometrisch eingeordnet und analysiert. Darüber hinaus werden drei bekannte Kompositionen, das „C-Dur Präludium“ von J. S. Bach, „Für Elise“ von L. v. Beethoven und „Eine kleine Nachtmusik“ von W. A. Mozart, im Symmetriemodell dargestellt und bewertet.

- Das Kapitel 11 stellt zwei Ansätze zur Tonarterkennung vor, die beide mit den Merkmalsvektoren des Symmetriemodells arbeiten. Der erste Ansatz erkennt Dur- und Molltonarten. Dies entspricht dem Stand der Technik. Der zweite Ansatz ist in der Lage, Kirchentonarten zu erkennen. Für die Evaluierung wird ein neuer Bewertungsmaßstab eingeführt.
- Im abschließenden Kapitel 12 werden der Chromavektor, der 6D-Vektor von Har- te, Sandler und Gasser und der Merkmalsvektor des Symmetriemodells mit den drei Maschinenlernverfahren *Hidden Markov Models* (HMM), *K Nearest Neighbours* (KNN) und dem neuen Verfahren, das HMM und KNN in einem Algorithmus vereint, kombiniert und für die Akkorderkennung verwendet. Es wird untersucht, welcher Merkmalsvektor in Verbindung mit welchem Maschinenlernverfahren für die Akkorderkennung am besten geeignet ist.
- Kapitel 13 fasst die wichtigsten Ergebnisse zusammen und gibt einen Ausblick, wie die hier erfolgten Arbeiten weitergeführt werden können.

Kapitel 2

Grundlagen der Tonalität

Die Aufgabe von geometrischen Tonräumen besteht darin, die Beziehungen von Tönen innerhalb von Musikstücken verständlich darzustellen, zu beschreiben und damit auch analysierbar zu machen. Das Ziel des vorliegenden Grundlagenkapitels besteht deshalb darin, die wichtigsten Beziehungen zwischen Tönen aufzuzeigen und in einen psychoakustischen und wahrnehmungspsychologischen Zusammenhang zu stellen. Diese Grundlagen stellen eine Basis dar, mit deren Hilfe die in Kapitel 3 beschriebenen Tonräume eingeordnet und bewertet werden können. Der Aufbau des Grundlagenkapitels ist wie folgt: Wir werden zunächst kurz auf den Begriff der „Tonalität“ eingehen, anschließend einen Überblick über das auditorische System des Menschen geben und dann die Grundlagen zur Wahrnehmung von Einzeltönen, Intervallen, Akkorden und Tonarten behandeln. Großen Einfluss auf die Wahrnehmung dieser Elemente hat der tonale Kontext, welcher Gegenstand des abschließenden Kapitels 2.7 ist.

2.1 Tonalität

Ein Musiksinal ist im Grunde genommen eine Aufeinanderfolge und Überlagerung von musikalischen Ereignissen. Diese Ereignisse stehen in ganz bestimmten Beziehungen zueinander und diese Beziehungen wiederum können auf verschiedenen Ebenen beschrieben werden. Eine wichtige Ebene hierbei ist der *Rhythmus*, welcher die zeitlichen Beziehungen zwischen musikalischen Ereignissen eines Musikstückes beschreibt. Eine weitere Ebene ist die *Klangfarbe*, welche das Frequenzspektrum und den zeitlichen Amplitudenverlauf der musikalischen Ereignisse zusammenfasst. Und die dritte Ebene ist die *Tonalität*, welche auch den Hauptgegenstand der vorliegenden Arbeit darstellt. Die Tonalität ist nach Piston als die „organisierte Beziehung von Tönen in Musik“ definiert [125]. Diese Beziehungen ergeben sich zum einen aus zeitlichen und zum anderen aus

tonhöhenmäßigen Relationen zwischen Tönen. Tonalität wiederum kann in die beiden Komponenten *Harmonie*, das ist der gleichzeitige oder aufeinanderfolgende Zusammenklang mehrerer Töne, und die Komponente *melodische Kontur*, das ist der im Großen und Ganzen wahrgenommene Tonhöhenverlauf eines Musikstückes, zerlegt werden. In der Musikgeschichte haben sich ganz unterschiedliche historisch und kulturell bedingte Tonalitätssysteme herausgebildet. Die vorliegende Arbeit konzentriert sich jedoch auf die Entwicklung von Modellen für eine spezielle Form der Tonalität, das ist die *Dur-Moll-Tonalität*, wie sie zwischen 1600 und 1900 entwickelt wurde [83, S. 3]. Diese spezielle Form der Tonalität – im Folgenden auch als „abendländische Tonalität“ bezeichnet – ist ein guter Ausgangspunkt für die Entwicklung von Tonraumsystemen. Denn sie stellt nach wie vor die Grundlage eines sehr großen Anteils aktueller Musik dar. Des Weiteren weist sie einen starken psychophysikalischen Bezug auf.¹ Damit ist vor allem gemeint, dass das abendländische Tonsystem zum einen am physikalischen Aufbau und der psychoakustischen Wirkung von Klängen² orientiert ist und sich zum anderen in seiner starken Hierarchisierung (vgl. [81]) an der Art und Weise der vermuteten neuronalen Tonsignalverarbeitung ausrichtet (vgl. [33]). Sie steht damit in einem Kontrast zur Atonalität, die hauptsächlich mit mathematisch-mengentheoretischen Verhältnissen von Tönen arbeitet [45]. Unter der Annahme, dass sowohl tonale als auch atonale Musik durch ähnliche oder gleiche kognitive Mechanismen verarbeitet werden, ist zu erwarten, dass viele Gesetze, die im Zusammenhang mit der Dur-Moll-Tonalität modelliert werden, auch auf die Klassifikation atonaler Musik oder anderer Tonalitätssysteme übertragen bzw. erweitert werden können.³ Dies ist jedoch nicht Gegenstand der vorliegenden Arbeit und soll deshalb an dieser Stelle nur als Ausblick stehen. Bevor wir im nächsten Kapitel zum auditorischen System des Menschen übergehen, sollen an dieser Stelle die wichtigsten Eigenschaften des abendländischen Tonsystems zusammengefasst werden:

- *Sonanz*: Hohe „Tonhaftigkeit“ und der Grad der Abwesenheit von Störungen, d.h. die Abwesenheit von Fluktuationen, Rauheiten und Schärfe[n] [42] sind ein wesentlicher Bestandteil der abendländischen Tonalität [159].

¹ Dies lässt sich z.B. an dem Gewicht erkennen, welches Autoren grundlegender Harmonielehrebücher der Betrachtung von physikalischen Schwingungen, Schwingungsverhältnissen und Obertönen sowie psychoakustischen Wahrnehmungsparametern wie Rauheit, Konsonanz und Dissonanz zumessen.

² Dazu gehören z.B. die Rauheit, die Schärfe, die Klanghaftigkeit und auch die Ähnlichkeit von Klängen (z.B. Oktav- und Quintähnlichkeit) [159].

³ Grund für diese Annahme liefert zum Beispiel Tuire Kuusi [87], die gezeigt hat, dass auch die Wahrnehmung atonaler Tonkombinationen stark durch psychoakustische Parameter wie sensorische Dissonanz geprägt ist.

- *Tonverwandtschaft*: Töne mit ähnlichen Obertonfrequenzen⁴, aber auch Akkorde mit gemeinsamen Tönen werden als verwandt betrachtet und treten oft zeitlich sehr dicht beieinander auf [159].
- *Grundtonbezogenheit*: Tonkombinationen werden vom Gehör auf die Existenz eines virtuellen Grundtones analysiert. Das abendländische Tonsystem arbeitet viel mit Klängen, die einen virtuellen Grundton erzeugen [159]. Der Verwandtschaftsgrad von Akkorden wird dadurch auch durch die Tonverwandtschaft der zugehörigen virtuellen Grundtöne bedingt /item *Hierarchisierung*: Die Wahrnehmung von Tönen erfolgt in Bezug auf einen zentralen Ton oder Akkord, der als Tonika bezeichnet wird. Die Tonika ist eine Art tonaler Referenzpunkt, auf den eine musikalische Passage ausgerichtet ist [81, S. 348]. Die Art der Hierarchie hängt vor allem von der gewählten Tonart ab.
- *Stabilität*: Je nach Ausprägung der zu einem Zeitpunkt etablierten Tonhierarchie kann jedem Ton eine Stabilität zugeordnet werden. Diese hängt von der Position des jeweiligen Tones in der tonalen Hierarchie ab und kann über die Zeitdauer eines Stückes variieren. Praktisch hängt die Stabilität der Töne von der Auftrittswahrscheinlichkeit und Länge sowie der metrischen Position der Töne ab [82]. Näheres dazu ist in Abschnitt 2.7.1 beschrieben.

2.2 Das auditorische System des Menschen

Das auditorische System des Menschen besteht aus der auditorischen Peripherie, das ist das Außen-, Mittel- und Innenohr, sowie dem zentralen Nervensystem. Das Außenohr filtert den empfangenen Schall richtungsbezogen und leitet diesen an das Trommelfell weiter. Das Mittelohr gibt den empfangenen Schall über die Gehörknöchelchen an das Innenohr weiter, wo die mechanischen Schwingungen in der Cochlea frequenzabhängig in neuronale Nervenaktionspotentiale umgewandelt und an die angeschlossenen Hörnervenfasern weitergegeben werden. Dadurch kommt es zu einer Kette neuronaler Erregungen längs der zentralen Hörbahn bis zum auditorischen Kortex [66]. Im Bereich der auditorischen Peripherie ist für die Wahrnehmung von Tonalität vor allem die Cochlea von Bedeutung. Diese zerlegt ein Schallsignal in einzelne Frequenzkomponenten. Bedingt durch den physiologischen Aufbau gelingt dies jedoch nur zu einem gewissen Grad. Sind z.B. die Frequenzen zweier gleichzeitig erklingender Sinustöne zu dicht

⁴ Das sind vor allem oktav-, quint- und terzverwandte Töne.

beieinander, ist der Mensch nicht in der Lage, zwei unabhängige Sinustöne wahrzunehmen, sondern er nimmt *einen* stark schwebenden oder durch Rauheiten geprägten Ton wahr. Ein Frequenzabschnitt, innerhalb dessen die Intensitäten zweier Sinustöne zu einer Gesamtintensität zusammengefasst werden und somit gemeinsam früher überschwellig werden als die Einzelsignale, wird als *Frequenzgruppe* bezeichnet [42], die spektrale Ausdehnung einer Frequenzgruppe als *Frequenzgruppenbreite*. In Abschnitt 2.4 wird gezeigt, dass die Frequenzgruppenbreite in einem engen Zusammenhang mit der Wahrnehmung von Intervallen und damit der Entwicklung des chromatischen Tonsystems steht.

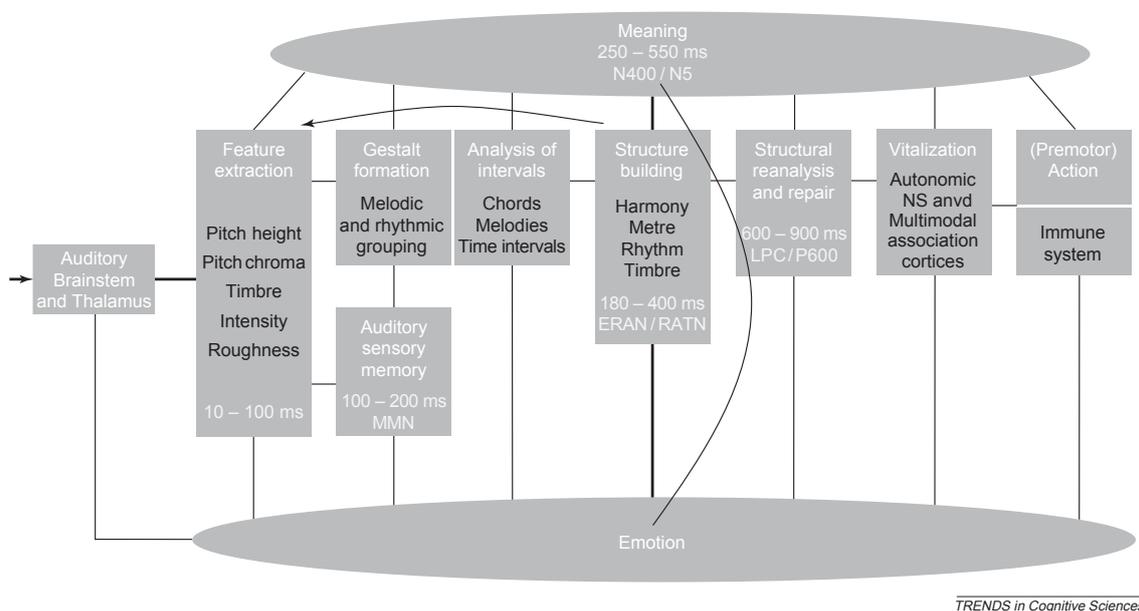


Abbildung 2.1 – Neurokognitives Modell der Musikwahrnehmung [79].

Die Wahrnehmung von Musik geht weit über Vorgänge im Innenohr hinaus. Entscheidende Verarbeitungsschritte erfolgen erst im menschlichen Gehirn. Eine tiefer gehende Betrachtung der menschlichen Musikwahrnehmung erfordert deshalb einen Blick auf die neuronale Verarbeitung von Musik. Ein solcher ist mit Abbildung 2.1 gegeben [79]. Bevor auf die einzelnen Blöcke des Modells eingegangen wird, sollen zunächst wichtige Hintergrundinformationen zu Kölschs Modell gegeben werden:

- Die Zeitangaben in den einzelnen Blöcken (10-100ms, 100-200ms etc.) bezeichnen den zeitlichen Abstand zwischen dem Auftreten eines musikalischen Ereignisses und einer zugehörigen Aktivitätsreaktion des Gehirns. Letztere wird z.B. mit Hilfe eines Elektroenzephalogramms (EEGs) gemessen und als *ereigniskorreliertes Potential (EKP)* bezeichnet.
- So bezeichnet z.B. die Abkürzung *N400* in [79] ein EKP, das durch ein negatives

Potential (N) gekennzeichnet ist und 400ms nach einem bestimmten musikalischen Ereignis auftritt. Die Abkürzung P600 bezeichnet ein EKP, das ein positives Potential (P) aufweist und 600ms nach einem musikalischen Ereignis auftritt.

- Die Abkürzungen MMN, ERAN/RATN und LPC bezeichnen besonders typische und häufig verwendete EKPs. Bei der MMN (Mismatch Negativity) handelt es sich um ein EKP, das ca. 150 bis 200ms nach Veränderung eines bestehenden Wahrnehmungstimulus hinsichtlich Frequenz, Dauer, Ort oder Intensität auftritt.
- Bei der *ERAN* (Early Right Anterior Negativity) handelt es sich um ein EKP, das ca. 180 bis 200ms in Experimenten mit isochronen, sich wiederholenden Stimulus gemessen werden kann [79]. Dies ist z.B. dann der Fall, wenn musiksyntaktisch irreguläre Akkorde auftreten [89].
- Das *RATN* (Right Anterior-Temporal Negativity) kann z.B. in Experimenten gemessen werden, in welchen das „Auftreten irregulärer Akkorde unvorhersagbar“ ist [79]. Die *RATN* ist durch eine größere Latenz sowie zeitliche Verteilung gekennzeichnet.
- Die *LPC* (Late Positive Component) bzw. P600 tritt ca. 600ms nach einem musikalischen Ereignis auf und „reflektiert die strukturelle Integration“ ... „einer regelbasierten sequentiellen Information“ [90].

Nachdem wichtige Abkürzungen aus Abbildung 2.1 erläutert wurden, soll nun auf die einzelnen Blöcke des Modells eingegangen werden: Die über die Hörnervenfasern übertragenen Aktivitätssignale werden zunächst vom auditorischen Gehirnstamm (Auditory Brainstem) und dem Thalamus empfangen und bereits auf verschiedene grundlegende Eigenschaften wie Tonhöhe, Klangfarbe, Rauheit, Intensität oder interaurale Differenzen analysiert. Dies erfolgt in einem Zeitraum von 10-100ms. Ziel hierbei ist z.B. die möglichst frühe Ermittlung auditorischer Gefahrensignale. Die ermittelten Informationen werden an das auditorische Gedächtnis (Auditory Sensory Memory) weitergegeben, wo ein Abgleich erfolgt und bei dem es im Fall von neuen Geräuschen nach ca. 100 bis 200ms zum Auftreten einer Mismatch Negativity (MMN) kommt. Nach Ermittlung grundlegender psychoakustischer Eigenschaften wird das Musiksignal einer „Gestaltformation“ unterzogen, d.h. anhand von Kriterien wie Ähnlichkeit, zeitlicher Nähe und Kontinuität gruppiert bzw. in getrennte Ströme zerlegt [14] und anschließend auf Akkorde und Melodien untersucht („Analysis of Intervals“). Die Aufgabe des Moduls „Structure Building“ besteht in der Identifikation syntaktischer Strukturen, ähnlich wie es bei Sprachsignalen der Fall ist. Die so ermittelten Informationen erlauben wiederum dem Modul „Structural

Reanalysis and Repair“ stark gestörte Signale zu reparieren bzw. zu vervollständigen. Der Block „Vitalization“ sorgt für die Auslösung von Emotionen oder anderen vitalen Reaktionen. Das Modul „(Premotor) Action“ schließlich ist für die Planung und Umsetzung musikbezogener körperlicher Bewegungen – wie z.B. Mitwippen – da. Alle bisher beschriebenen Blöcke sind mit den Modulen „Meaning“ und „Emotionen“ verbunden, welche für die Verknüpfung der auditorischen Informationen mit bedeutungsbehafteten Begriffen wie z.B. „hell“ oder „rauh“ sowie für die Ausgabe von Emotionen sorgen.

2.3 Einzeltöne

Die in der realen Musik vorkommenden Töne werden als *komplexe Töne* bezeichnet. Ihre Signalform weicht deutlich von einfachen sinusförmigen Tönen ab und kann durch eine Überlagerung einer endlichen Zahl von sinusförmigen *Teiltönen*, auch als *Obertöne* oder *Harmonische* bezeichnet, beschrieben werden [159].⁵ Für die Analyse der Tonalität aus Audiosignalen haben die folgenden Eigenschaften von Einzeltönen Bedeutung.

2.3.1 Das Obertonspektrum

Anzahl, Frequenz und Energie der Obertöne eines Klanges haben großen Einfluss auf die Klangfarbe, welche zur Bestimmung der Tonalität von Musikstücken jedoch von untergeordneter Bedeutung ist. Trotzdem hat das Obertonspektrum eine große Bedeutung für das abendländische Tonsystem. William Sethares [149] zeigte, dass die in bestimmten Kulturen vorhandenen Tonsysteme und Skalen auf Basis der Obertonspektren der verwendeten Musikinstrumente erklärt werden können. Ursache hierfür sind vor allem klangfarbenbedingte Unterschiede in den Intervalldissonanzen (Abschnitt 2.4.2). Dies wird im Zusammenhang mit Intervalldissonanzen (Abschnitt 2.4) und der Herleitung der Diatonik (Abschnitt 2.6.1) noch einmal aufgegriffen.

2.3.2 Tonhöhe

Die Tonhöhe ist vor allem für die Analyse der melodischen Kontur bzw. für die Separation polyphoner Musiksignale in Einzelstimmen von Bedeutung. Die Tonhöhe von komplexen Tönen ist aufgrund der Überlagerung vieler Teiltöne nicht eindeutig bestimmbar und

⁵ Die komplexen Töne teilen sich in die Gruppen harmonische komplexe Töne, angenähert harmonische komplexe und geringharmonische Töne ein, wobei sich diese Kategorien auf den Grad der Abweichung der Teiltonfrequenzen von den ganzzahligen Vielfachen der Grundfrequenzen bezieht.

muss durch Einbezug des gesamten Spektrums rekonstruiert werden⁶. Durch einen Mustervervollständigungsprozess ist das auditorische System des Menschen darüber hinaus in der Lage, die fehlenden Grundfrequenzen in unvollständigen Spektralmustern zu rekonstruieren und wahrnehmbar zu machen. Solche nicht im Signal vorhandenen, aber wahrnehmbaren Grundfrequenzen werden als *virtuelle Grundtöne* bezeichnet [159].

2.3.3 Oktavähnlichkeit, Tonigkeit und Oktavlage

Klänge, deren Grundfrequenzen im Abstand einer oder mehrerer Oktaven ($f_2 = f_1 2^n$) sind, werden als sehr ähnlich wahrgenommen. Man spricht hier auch von *Oktaväquivalenz*. Diese Ähnlichkeit kann sowohl durch das einfache Frequenzverhältnis der beiden Klänge als auch auf Basis der übereinstimmenden Obertonspektren erklärt werden. Oktavverwandte Klänge besitzen dadurch auch einen sehr hohen *Verschmelzungsgrad* [155]. Die Gemeinsamkeit zweier oktavähnlicher Töne wird als *Tonigkeit* bezeichnet. So besitzen alle Töne mit den Frequenzen $f_0 * 2^n$ die gleiche Tonigkeit, aber unterschiedliche *Oktavlagen*. Die Tonigkeit ist besonders für die Analyse der Harmonie von Bedeutung. Hier spielt die Oktavlage von Klängen nur eine untergeordnete Rolle. Interessant in diesem Zusammenhang sind Ergebnisse von J.D. Warren [165], der zeigte, dass Tonigkeit und Oktavlage in unterschiedlichen Gehirnregionen verarbeitet werden. Abbildung 2.2a zeigt ein Modell von Wilhelm Moritz Drobisch [38], das den Zusammenhang zwischen Tonigkeit und Oktavlage geometrisch veranschaulicht.⁷ Bei der Analyse von Audiosignalen findet sich die Bedeutung der Tonigkeit in der Verwendung von Chromavektoren wieder.

2.3.4 Quintähnlichkeit und Quintverwandtschaft

Klänge, deren Tonigkeiten im Abstand einer Quinte ($f_2 = f_1 \frac{3}{2} * 2^n$) stehen, werden als quintverwandt bezeichnet. Wie bei oktavähnlichen Klängen stehen die Grundfrequenzen quintverwandter Klänge in einfachen Frequenzverhältnissen $\frac{3}{2}$. Auch fallen die Obertonspektren von quintverwandten Klängen weitgehend zusammen, was zu einer hohen sensorischen Konsonanz führt. Darüber hinaus tragen quintverwandte Klänge häufig zur Ausbildung virtueller Grundtöne bei [159]. Abbildung 2.2b zeigt eine Erweiterung von Drobischs Geometriemodell, bei dem Roger Shepard [152] das Modell um zwei zusätzliche die Quintähnlichkeit repräsentierende Dimensionen ergänzt. Zusammenfassungen

⁶ Um dieses Problem zu lösen, führt Ernst Terhardt [159] den Begriff der Spektraltonhöhen ein. Spektraltonhöhen sind alle Tonhöhen, die in einem Klang theoretisch wahrgenommen werden können. Spektraltonhöhen wiederum werden zu virtuellen Tonhöhen kombiniert.

⁷ Drobischs Modell ist die Grundlage für den in Abschnitt 4.3.4 vorgeschlagenen Tonigkeits-Tonhöhenraum.

und weitere Informationen zu Shepards Modell können z.B. in [80, S. 347] oder [81] gefunden werden.

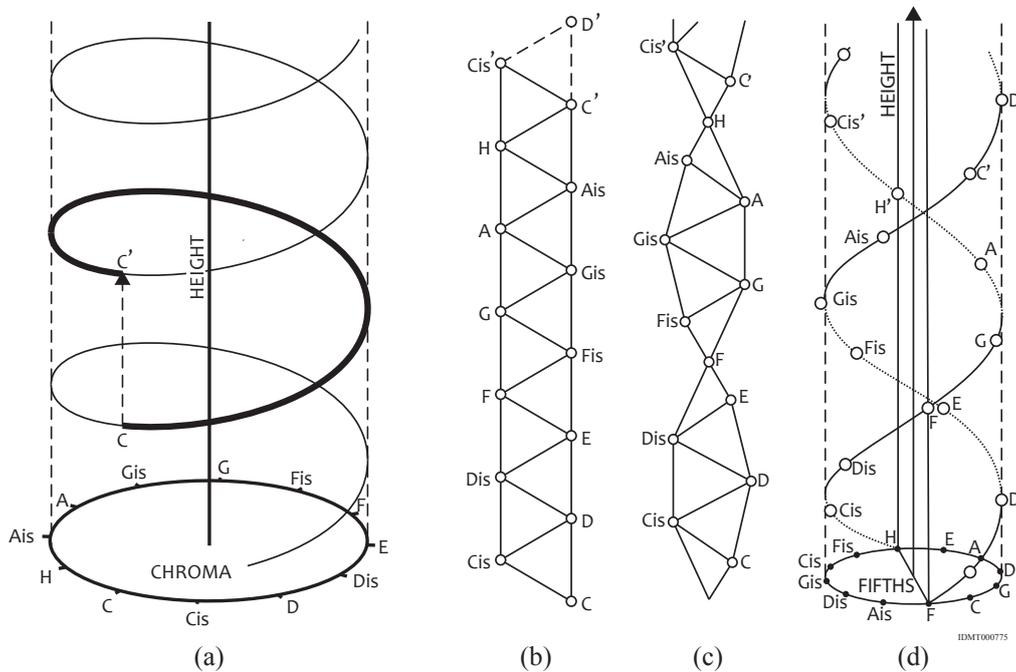


Abbildung 2.2 – (a) Drobischs geometrisches Modell zur Veranschaulichung der Begriffe Tonigkeit und Oktavlage: Der in der XY-Ebene dargestellte Kreis repräsentiert die Zirkularität der Tonigkeit, die Z-Achse die Tonhöhe. (b), (c), (d) Roger Shepard erweiterte Drobischs Modell um zwei weitere Dimensionen, welche die Quintverwandtschaft von Tonigkeiten repräsentieren sollen (aus [152]).

2.3.5 Identifikation bzw. Benennung von Tönen und Tonigkeiten

Die Identifikation von Tönen erfolgt in dieser Arbeit durch Verwendung eines *Halbtonindex* n und durch Verwendung von *Tonigkeitsnamen*. Eine temperierte Stimmung (Abschnitt 2.6.3) vorausgesetzt ergibt sich der Halbtonindex n_c eines Tones mit der Frequenz f mit $n_c = 12 \cdot \text{ld}(f/16,35\text{Hz})$. Die Referenzfrequenz 16,35Hz ist die Frequenz des Subkontra-Cs [159, S. 204], welches dadurch den Halbtonindex $n = 0$ erhält. Dadurch ergibt sich die in Tabelle 2.1 dargestellte Zuordnung von Halbtonindizes n_c zu Tonigkeitsnamen⁸ TN (NN). Weitere Details zur Namensgebung und Indexzuordnung können z.B. in [112] oder [84] gefunden werden.

⁸ Wir beschränken uns in dieser Arbeit auf die Angabe von Tonigkeitsnamen. Die Musiktheorie vergibt darüber hinaus noch unterschiedliche Namen für die unterschiedlichen Oktavlagen einer Tonigkeit (z.B. Großes C,

n_c	0	1	2	3	4	5	6	7	8	9	10	11	12	13	...
TN	c	cis/des	d	dis/es	e	f	fis/ges	g	gis/as	a	ais/b	h	c	cis/des	...

Tabelle 2.1 – Identifikation von Tönen in Form von Tonindizes n_c und Tonigkeitsnamen

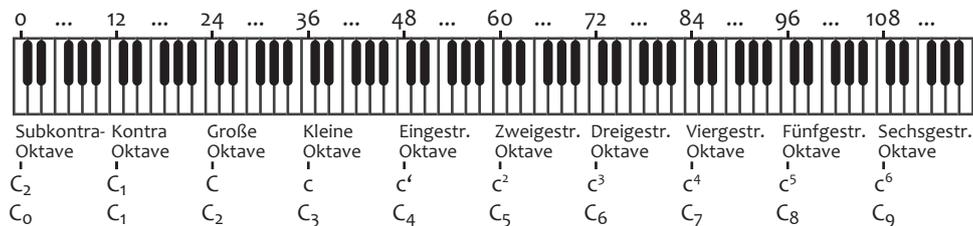


Abbildung 2.3 – “Die chromatische Tonskala der abendländischen Musik und die Arten der Tonbezeichnung. Obere Reihe: Tonindex. Unter dem Manual: Europäisch traditionelle Tonbezeichnung. Unterste Reihe: Bevorzugte internationale Tonbezeichnung“ [159, S. 202]

2.4 Intervalle

Die zeitlich sequentielle oder simultane Kombination aus zwei Tönen wird als Intervall bezeichnet.

2.4.1 Identifikation bzw. Benennung von Intervallen

Die Identifikation von Intervallen erfolgt über Frequenzverhältnisse, Intervallbezeichnungen und Halbtonabstände. Das *Frequenzverhältnis* FV zweier Töne ergibt sich mit $FV = f_2/f_1$. Dabei ist f_1 die Grundfrequenz des ersten und f_2 die Grundfrequenz des zweiten Tones. Der *Halbtonabstand* HTA ergibt sich mit $HTA = |n_{c_2} - n_{c_1}|$, wobei n_{c_2} der Halbtonindex des ersten und n_{c_1} der Halbtonindex des zweiten Tones ist (Abschnitt 2.3.5). Die Zuordnung von *Intervallbezeichnungen*⁹ zu Frequenzverhältnissen¹⁰ und Halbtonabständen ist in Tabelle 2.2 dargestellt [112].

Bei der Verwendung von Intervallbezeichnungen ist im Zusammenhang mit chroma-

kleines C, eingestrichenes C usw.).

⁹ In der musiktheoretischen Literatur [84, S. 5] wird über die in Tabelle 2.2 angegebenen Intervalle hinaus noch zwischen übermäßigen und verminderten Intervallen unterschieden. Das wichtigste Beispiel hier ist der Tritonus, welcher als übermäßige Quarte oder als verminderte Quinte gedeutet wird.

¹⁰ Bei den angegebenen Frequenzverhältnissen handelt es sich um Annäherungen. Die tatsächlichen Frequenzverhältnisse hängen von der jeweiligen Instrumentenstimmung ab (Abschnitt 2.6.3). Auf eine Angabe der Frequenzverhältnisse für die kleine und große Sekunde sowie den Tritonus wurde verzichtet, da die zugrundeliegenden Frequenzverhältnisse in der Praxis sich aus den eingestimmten Frequenzen der restlichen Töne ergeben.

IVB	Prime	Kleine Sekunde/ Halbton	Große Sekunde/ Ganzton	Kleine Terz	Große Terz	Quarte	Tritonus/ Verminderte Quinte/ Übermäßige Quarte	Quinte	Kleine Sexte	Große Sexte	Kleine Septime	Große Septime	Oktave
HTA	0	1	2	3	4	5	6	7	8	9	10	11	12
FV	1:1			6:5	5:4	4:3		3:2	8:5	5:3	16:9	15/8	2:1

Tabelle 2.2 – Zuordnung von Halbtonabständen HTA, Frequenzverhältnissen FV und Intervallbezeichnungen IVB

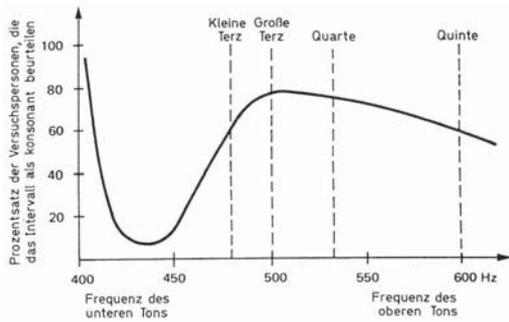
vektorbasierter Audioanalyse (siehe Abschnitt 7.2) zu beachten, dass hier nicht zwischen sogenannten *Komplementärintervallen* [84, S. 7], [112] unterschieden werden kann. Komplementärintervalle sind Intervalle, die sich gegenseitig zur Oktave ergänzen. Dazu gehören z.B. Quarte und Quinte, große Terz und kleine Sexte. Ursache hierfür ist die Tatsache, dass der Chromavektor Tonigkeiten repräsentiert und damit keine Tonhöheninformation bereithält. Es ist damit z.B. keine Unterscheidung zwischen der Quinte $c-g$ oder der Quarte $g-c$ möglich.

2.4.2 Intervallkonsonanz und -dissonanz

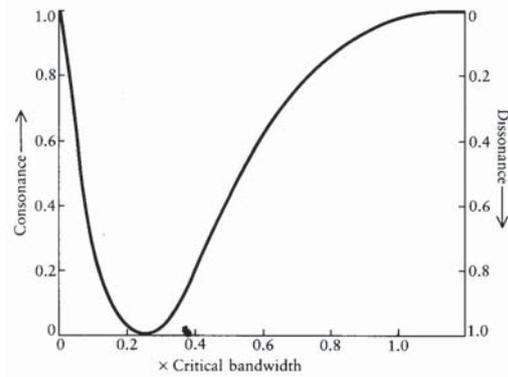
Die Wahrnehmung von Intervallen ist hauptsächlich durch Konsonanzen bzw. Dissonanzen gekennzeichnet. Verschiedene Maße der *Intervallkonsonanz* wurden von Euler [41] (Einfachheit der Frequenzverhältnisse), Helmholtz [67] (Rauheit), Stumpf [155] (Tonverschmelzung), Malmberg¹¹ [106], Plomp und Levelt [126], Hutchinson und Knopoff [70] (Frequenzgruppenbreite), Kameoka und Kuriyagawa [76] (Dissonanzwerte für komplexe Töne mit unterschiedlich vielen Obertönen), Terhardt [159] (Eindeutigkeit des virtuellen Grundtones als Konsonanzmaß), Sethares [149] und weiteren erstellt und ausführlich diskutiert (z.B. [82, S. 55ff], [149], [27]). Insgesamt kann nach heutigem Erkenntnisstand gesagt werden, dass die Wahrnehmung von Intervallen in einem engen Zusammenhang mit der cochlearen Frequenzgruppenbreite steht (Abschnitt 2.2, Abbildungen 2.4a, 2.4b und 2.4c). Die besondere Bedeutung der musikalisch sehr bedeutsamen Intervalle (Quinte, Quarte, Terz und Sexte) kann zum einen physiologisch dahingehend erklärt werden, dass obertonreichen Klänge in diesen Intervallen eine sehr geringe Dissonanz aufweisen (Abbildung 2.4). Zum anderen gibt es neurophysiologische Erklärungen, die besagen,

¹¹ Malmberg untersuchte die Wahrnehmung von Intervallen mit musikalisch trainierten Personen.

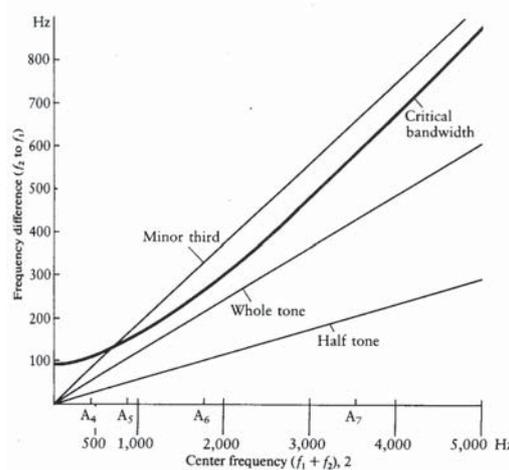
dass Intervalle mit einfachen Frequenzverhältnissen neuronale Feuermuster mit ebenfalls einfachen Frequenzverhältnissen erzeugen (siehe [82, S. 52]). Die kognitive Last bei der Wahrnehmung von Intervallen mit einfachen Frequenzverhältnissen könnte dadurch geringer sein, als die anderer.



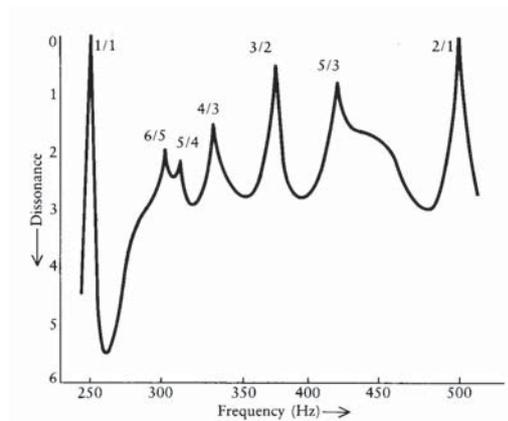
(a) Die wahrgenommene Konsonanz von Intervallen aus sinusförmigen Tönen (aus [109])



(b) Frequenzgruppenbreite und Intervallkonsonanz (aus [124])



(c) Frequenzgruppenbreite und das Intervall der kleinen Terz (aus [124])



(d) Sensorische Konsonanz von Intervallen aus komplexen Tönen mit sechs Obertönen (aus [124])

Abbildung 2.4 – Zusammenhang zwischen wahrgenommener Intervallkonsonanz und Frequenzgruppenbreite

2.4.3 Grenzen der Intervallkonsonanz zur Erklärung der Harmonik

Sobald drei oder mehr Töne gleichzeitig erklingen, kann der wahrgenommene Wohlklang nur noch begrenzt auf Basis der Intervalldissonanz vorherbestimmt werden. So versagen

auf Intervalldissonanz basierende Verfahren bei der korrekten Berechnung des Wohlklanges bestimmter Umkehrungen von Dur- und Mollakkorden und der Dissonanz des übermäßigen Dreiklanges [27]. Auch die geringere Stabilität von Mollakkorden gegenüber Durakkorden (Abschnitt 2.7.2) [82, S. 50] und der gegenteilige emotionale Charakter von Dur- und Mollakkorden [28], [29], [30], [77], [31] sind auf Basis von Intervalldissonanzen nur sehr schwer erklärbar [132], [27]. Lösungsansätze sind hier z.B. der Einbezug der Auditory Scene Analysis¹², kognitiver Effekte wie Akkordspannungen [27] oder hierarchischer Strukturierung von Musiksignalen durch das auditorische System des Menschen [80], [81], [33].

2.4.4 Satztechnische Bedeutung von Intervallen

Satztechnisch haben die verschiedenen Intervalle in ganz unterschiedlichen Bereichen Bedeutung, welche hier kurz zusammengefasst werden soll:

- *Prime sowie kleine und große Sekunde* haben vor allem für die Führung einzelner Stimmen Bedeutung. Nach dem „Gesetz des nächsten Weges“ soll der Komponist versuchen, „in jeder Stimme den nächst erreichbaren Akkordton auf kürzestem Wege zu finden“. In den Mittelstimmen ist sogar das „Liegenlassen gemeinsamer Akkordtöne“ [84, S. 31] und damit die Verwendung der Prime ideal. Eine konsequente Anwendung dieser Regeln führt dazu, dass *zeitlich benachbarte Tonigkeiten innerhalb einer Stimme* in der Mehrzahl eine Prime, kleine oder große Sekunde haben.
- *Kleine und große Terz*: Die harmonischen Grundpfeiler der tonalen Musik sind der Dur- sowie der Mollakkord, welche als die einzigen in sich ruhenden Dreiklänge betrachtet werden [85, S. 142]. Diese Tonkombinationen bestehen aus geschichteten großen und kleinen Terzen, was dazu führt, dass *gleichzeitige Tonigkeiten in benachbarten Stimmen* vorrangig Intervalle von einer kleinen oder großen Terz haben.¹³
- *Quinte/Quarte*: Das Intervall der Quinte hat hauptsächlich aus der Perspektive der Kadenz Bedeutung: In musikalischen Musikstücken werden Akkorde zeitlich so angeordnet, dass eine Tonart bzw. ein Tongeschlecht etabliert wird und das Gefühl der „Geschlossenheit“ einer Klangfolge [84, S. 11] entsteht. Dies wird vor allem

¹² Albert S. Bregman [14, S. 215] berichtet von Experimenten, bei denen festgestellt wurde, dass die Wahrnehmung von Intervalldissonanzen, Schwebungen etc. komplett unterdrückt wurden, wenn die beiden Töne in unterschiedlichen auditorischen Strömen wahrgenommen wurden.

¹³ Die kleine Terz und die große Terz haben zusätzlich noch im Zusammenhang mit den Nebendreiklängen eine große Bedeutung bei der Benachbarung von Akkorden.

durch den Einsatz der drei quintverwandten Funktionen Tonika, Subdominante und Dominante erreicht (siehe Abschnitt 5.3.1 „Funktionen“). Dies wiederum bedeutet, dass *zeitlich benachbarte Akkorde* innerhalb von Musikstücken hauptsächlich dem Intervall der Quinte folgen.

- Der *Tritonus* hat eine entscheidende Bedeutung für die Definition der Tonart: Wie später noch gezeigt werden wird (Abschnitt 5.4.1), bildet der Tritonus den Anfang und das Ende des von einem diatonischen Tonset im tonartübergreifenden Quintenzirkel *F* belegten Halbkreises (Abbildung 5.6) und weist ein starkes Auflösungsbestreben in Richtung eines tonalen Zentrums auf (Abschnitt 5.4.1). Dies ist auch einer der Gründe, warum der Tritonus Bestandteil von Zwischendominanten ist (Abschnitt 5.4.2) und in Kombination mit weiteren Tönen häufig als Modulationsmittel auftritt (Abschnitt 5.5.3).

2.5 Akkorde

Eine Kombination aus drei oder mehr Tönen wird als Akkord bezeichnet. Hier sollen nun die wichtigsten Akkordtypen vorgestellt werden. Die Klassifikation von Akkorden kann einmal nach psychoakustischen perzeptuellen Kriterien und ein weiteres Mal nach rein mathematisch mengentheoretischen Kriterien erfolgen.¹⁴

2.5.1 Benennung von Dur- und Molldreiklängen

Im abendländischen Tonsystem treten zwei Akkorde besonders stark hervor, das sind der Dur- und der Molldreiklang. Diese beiden Akkorde dienen als Referenz, auf welche alle anderen Akkorde bezogen werden. Zur Benennung von Dur- und Molldreiklängen hat die Musiktheorie ganz unterschiedliche Systeme hervorgebracht. In der vorliegenden Arbeit werden die folgenden verwendet:

- Die Angabe von *Grundtönen* (Akkordsymbole): Der tiefste Ton eines Dur- oder Molldreiklanges in Grundstellung (siehe unten) wird als *Grundton* bezeichnet.¹⁵ Das *Tongeschlecht* des Akkordes wird in Form der *Groß- und Kleinschreibung* des

¹⁴ Ein Überblick über mengentheoretische Klassifizierungsverfahren sowie ein Vergleich mit der tatsächlichen Wahrnehmung wurden z.B. von Tuire Kuusi [87], [88] gegeben. Weitere Arbeiten in diesem Zusammenhang sind bei Arthur G. Samplaski [138] und Cheryl L. Bruner [16] zu finden.

¹⁵ Der Grundton eines Akkordes ist unabhängig von der Umkehrung. Er darf nicht mit dem tiefsten Ton verwechselt werden. So ist der Grundton des Dreiklanges *e-g-c* der Ton *c*, der tiefste Ton jedoch der Ton *e*.

Akkordes gekennzeichnet. Der Akkord C-Dur wird somit einfach über Angabe des Großbuchstabens „C“ und der Akkord c-Moll über Angabe des Kleinbuchstabens „c“ bezeichnet.

- Die Angabe von *Tonigkeitsbezeichnungen* bzw. *Halbtonindizes*: Entsprechend Tabelle 2.1 können Akkorde über die Angabe der Akkordtöne identifiziert werden. So kann über die Angabe der Buchstaben „c-e-g“ bzw. der Halbtonindizes „0-4-7“ der C-Dur-Dreiklang beschrieben werden.
- Die Angabe von *Intervallschichtungen*: Entsprechend den in Abschnitt 2.4.1 eingeführten Intervallbezeichnungen kann ein Akkord auch über die Schichtung von Intervallen identifiziert werden. Der Dreiklang „c-e-g“ kann damit auch in Form der Angabe von „Große Terz + Kleine Terz“ bzw. der Halbtonabstände „4-3“ beschrieben werden. Die Verwendung von Intervallschichtung ist vor allem dann von Bedeutung, wenn nur die Verhältnisse der einzelnen Töne des Akkordes zueinander eine Rolle spielen, nicht jedoch die genaue Tonhöhe.
- Die Angabe von *Akkordumkehrungen*: Mit Hilfe von Akkordumkehrungen wird die *tonhöhenmäßige Reihenfolge* der Töne von Dur- und Mollakkorden beschrieben. Im Beispiel des C-Dur-Dreiklanges wird die Reihenfolge c-e-g (Terz-Terz) als *Grundstellung*, die Reihenfolge e-g-c (Terz-Quarte) als *1. Umkehrung* und die Reihenfolge g-c-e (Quarte-Terz) als *2. Umkehrung* bezeichnet.¹⁶
- Die Angabe von *Generalbassbezeichnungen*: Die Generalbassbezeichnung von Akkorden ähnelt sehr der Angabe von Intervallschichtungen. Während bei der Intervallschichtung die *Intervalle benachbarter Töne* angegeben werden, werden bei der Generalbassbezeichnung die *Intervalle bezogen auf den tiefsten Ton* des Akkordes bezeichnet. So wird die 1. Umkehrung eines Akkordes (z.B. e-c-g) als Terzsextakkord bzw. als Sextakkord¹⁷ bezeichnet und die 2. Umkehrung eines Akkordes (z.B. g-c-e) als Quartsextakkord.
- Angabe von *Funktionsbezeichnungen*: In einer gegebenen Tonart besitzen Akkorde eine sogenannte Funktion (z.B. Tonika, Dominante oder Subdominante etc.). Dies ermöglicht es auch, Akkorde über die Angabe von Funktionsbezeichnungen zu identifizieren. Weiteres dazu ist in Abschnitt 5.3.1 ausgeführt.

¹⁶ Spielt die relative Tonhöhe von Akkordtonigkeiten eine Rolle, so werden die Akkordtonigkeiten in Reihenfolge ansteigender Tonhöhe angegeben. Im Beispiel des Akkordes „e-g-c“ (1. Umkehrung des C-Dur-Akkordes) wird dabei davon ausgegangen, dass e der tiefste, g der mittlere und c der höchste Ton ist.

¹⁷ Die „Terz“ bezieht sich auf die Töne e-g und die Sexte auf das Intervall e-c.

- *Kombinationen* aus den vorherigen Angaben: In vielen Fällen werden auch die zuvor beschriebenen Akkordbezeichnungen kombiniert. Ein Beispiel ist der Dominantseptakkord. Diese Bezeichnung ist eine Kombination aus *Funktionsbezeichnung* (Dominante) und *Generalbassbezeichnung* (Septakkord).

Eine in dieser Arbeit nicht verwendete, aber in der Literatur häufig vorkommende Art der Akkordidentifikation ist die Benennung von Akkorden über Tonleiterstufen.¹⁸ Damit wird ein Akkord über die Position seines Grundtones auf einer gerade verwendeten Tonleiter benannt. Die Positionsangabe erfolgt in römischen Zahlen, *I, II, III* (Durakkorde) oder *i, ii, iii* (Molldreiklänge) etc. Die Angabe des Tongeschlechts erfolgt über Groß- und Kleinschreibung der Zahlensymbole.

2.5.2 Akkordtypen

An dieser Stelle müssten jetzt eigentlich die in dieser Arbeit verwendeten Akkordtypen beschrieben werden. Jedoch soll hier auf das Kapitel „Repräsentation musikalischer Grundelemente im Symmetriemodell“ verwiesen werden (Kapitel 5). In diesem Kapitel werden verschiedene in der klassischen Komposition verwendete Akkorde im Kontext des Symmetriemodells diskutiert. Der Autor entschied sich für diesen Schritt, da das Symmetriemodell viele Eigenschaften von Akkorden geometrisch veranschaulicht und diese dadurch womöglich einfacher zugänglich macht.

2.5.3 Wahrnehmung von Akkorden

Wie bereits im Abschnitt 2.4.3 angedeutet ist die Wahrnehmung von Akkorden neben der Wahrnehmung von Intervallkonsonanzen und -dissonanzen durch weitere Wahrnehmungsphänomene gekennzeichnet. Hier soll nun ein Überblick über wichtige Eckpunkte der Akkordwahrnehmung gegeben werden:

- *Intervallkonsonanz und -dissonanz*, welche in Abschnitt 2.4.2 bereits näher erläutert wurden, spielen bei der Wahrnehmung von Drei- und Mehrklängen eine wichtige Rolle [87]. Die höchste Konsonanz weist der Durdreiklang auf. Zum einen bilden die Töne dieses Dreiklanges Intervalle mit einfachen Frequenzverhältnissen. Zum anderen kommt das durch den Durdreiklang gebildete Spektralmuster dem spektralen Muster eines obertonreichen Tones sehr nahe. Ähnliches gilt für

¹⁸ Gründe für eine Nichtverwendung von Stufenangaben sind z.B. in [47].

den Molldreiklang, wobei hier bereits die kleine Terz nicht mehr einem regulären Obertonmuster entspricht.

- *Spannung/Auflösung*: Es gibt Akkorde, die ein starkes Bestreben haben, „sich in andere Akkorde aufzulösen“. Dieses Bestreben wird als „*Auflösungsbestreben*“ bezeichnet. Mit Auflösung ist laut [85] die „*zwingende Weiterführung eines dissonierenden Intervalls in ein konsonierendes*“ gemeint. Ein sogenannter „*spannungsgeladener*“ Akkord wird von einem „*aufgelösten*“ Akkord gefolgt. Das Auflösungsbestreben eines Akkordes kann sich zum einen aus dem musikalischen Kontext ergeben (Abschnitt 2.7). Zum anderen gibt es Klänge, die in sich, d.h. ohne Existenz weiterer Akkorde, ein starkes Auflösungsbestreben aufweisen (siehe Abschnitt 5.3.3). Norman D. Cook [27] zeigte, dass rein intervalldissonanzbasierte Ansätze nicht in der Lage sind, das Auflösungsbestreben von Dreitonkombinationen vorherzusagen.
- *Wahrgenommene Verwandtschaft bzw. Nähe (Proximity)*: Die wahrgenommene Nähe zweier beliebiger Akkorde ergibt sich laut Kuusi [87] hauptsächlich aus dem Grad der Konsonanz, der Ähnlichkeit zu vertrauten tonalen Akkorden sowie der Nähe zur *Ganztonreihe*¹⁹. Weitere Faktoren sind laut Kuusi die *Oktavlage*²⁰ (Register) und *Oktavbreite*²¹ (Width) bzw. *Lage*. Die Nähe von Dur- und Molldreiklängen ergibt sich durch die Zugehörigkeit der Akkordtöne zu einer gemeinsamen Tonleiter (*leitereigene Klänge*), die Quintverwandtschaft der Grundtöne (Abschnitt 2.3.4), die Anzahl der gemeinsamen Töne und den musikalischen Kontext, in welchem die Akkorde erklingen [101, S. 53ff], [83] (Abschnitt 2.7.1). Die Verwandtschaft von Dur- und Mollakkorden wird im Zusammenhang mit dem Symmetriemodell noch einmal in Abschnitt 5.3.4 ausführlicher behandelt.
- *Affekt bzw. emotionaler Charakter*: Im Gegensatz zu Intervallen²² besitzen Tonkom-

¹⁹ Die Ganztonreihe ist eine Folge von Tönen, wobei jeweils zwei aufeinander folgende Töne das Intervall einer großen Sekunde bzw. eines Ganztones (Abschnitt 2.4.1) besitzen.

²⁰ Allgemein gesagt beschreibt die Oktavlage eines Akkordes den ungefähren Frequenzbereich, in welchem sich die Grundfrequenzen des Akkordes befinden. Musiktheoretisch gesehen regelt die Lage entsprechend [85, S. 147] „zum einen im vierstimmigen Satz das Verhältnis der drei Oberstimmen zueinander“. Je nachdem, wie dicht die Stimmen beieinander liegen, redet man von *enger Lage* oder *weiter Lage*.

²¹ Ein Akkord muss nicht zwangsweise nur aus Tönen einer Oktave zusammengesetzt werden, sondern kann auch aus Tönen mehrerer Oktaven bestehen. Der Begriff „Oktavbreite“ charakterisiert sehr grob die Breite des Frequenzbereiches, in welchem sich ein Akkord befindet.

²² Maher [105] untersuchte, ob bereits simultan präsentierte musikalische Intervalle unterschiedliche psychologische Effekte haben. Die Ergebnisse konnten keinen klaren Zusammenhang zwischen Emotion und Intervall herausstellen.

binationen aus drei oder mehr Tönen zum Teil einen sehr deutlichen emotionalen Charakter. Der Positiv-Negativ-Charakter von Dur- und Molldreiklängen wurde unter anderem von Robert G. Crowder ausführlich untersucht, welcher feststellte, dass bereits dreijährige Kinder die klassischen Stereotypen „Dur-Positiv“ und „Moll-Negativ“ wahrnehmen [28], [29], [30], [77], [31]. Ein sehr ausführlicher Überblick über den emotionalen Charakter von Akkorden und weiteren musikalischen Elementen ist in der Dissertation von Emery Schubert [142] zu finden.

2.6 Tonsets, Tonleitern, Modi und Skalen

Nachdem ein grober Überblick über Einzeltöne, Intervalle und Akkorde gegeben wurde, soll nun die nächsthöhere tonale Organisationsebene besprochen werden, das ist die Tonart. Wir unterscheiden in diesem Zusammenhang die Begriffe „Tonset“, „Tonleiter“, „Modus“ und „Skale“. Das *Tonset* ist eine ungeordnete Auswahl von Tonigkeiten²³, die *vorwiegend* in einem musikalischen Abschnitt verwendet werden. Sie stellt somit das tonale Grundmaterial eines Musikstückes bereit.²⁴ Der wahrgenommene Charakter eines Musikstückes hängt jedoch nicht allein von den verwendeten Tönen, sondern auch stark von der Reihenfolge und einem gewählten Anfangs- und Endton ab, auf den bezogen alle Töne wahrgenommen werden (siehe Abschnitt 2.7). Eine der Tonhöhe bzw. der Tonigkeit entsprechend *geordnetes Tonset* zusammen mit einem gewählten Anfangs- und Endton wird in dieser Arbeit als *Tonleiter* bezeichnet. Die Begriffe *Modus* und *Skale* sind Synonyme für den Begriff *Tonleiter*.

2.6.1 Das diatonische Tonset

Das diatonische Tonset²⁵ ist die in der abendländischen Musik am häufigsten verwendete Auswahl von Tönen. So bilden z.B. die weißen Tasten auf dem Klavier, d.h. die Tonigkeiten *c, d, e, f, g, a* und *h* ein diatonisches Tonset. Ordnet man das diatonische Tonset in einer der Tonhöhe entsprechenden Reihenfolge, so erhält man eine Reihe aus fünf Ganz- und zwei Halbtonschritten. Dieser „spezifische Wechsel von Ganz- und Halbtonstufen“

²³ Um eine einfache Lesbarkeit und Nähe zu existierender Literatur zu ermöglichen, reden wir in dieser Arbeit von „*Tonset*“, obwohl es sich genaugenommen um ein „*Tonigkeitsset*“ handelt.

²⁴ Synonyme für die Begriffe „*Tonset*“ und „*Tonleiter*“ sind auch die Begriffe „*Materialtonleiter*“ und „*Gebrauchstonleiter*“ [112, S. 85].

²⁵ Wir reden hier der Einfachheit halber von „dem diatonischen Tonset“. Genau genommen kann auf jeder existierenden Tonigkeit ein diatonisches Tonset aufgebaut werden.

wird als „*Diatonik*“ bezeichnet [112, S. 85]. Es gibt verschiedene Möglichkeiten, die sieben Tonigkeiten eines diatonischen Tonsets herzuleiten: Die pythagoreische Herleitung besteht darin, ausgehend von einer Referenztonigkeit drei Quinten aufsteigend und drei Quinten absteigend zu stimmen (siehe auch Abschnitt 4.2.3). Auch kann das diatonische Tonset über Intervalldissonanzen erklärt werden: In Abschnitt 2.4.2, Abbildung 2.4 wurde gezeigt, dass obertonreiche Klänge besonders konsonant sind, wenn deren Grundfrequenzen in den Verhältnissen 1:1, 1:2, 2:3, 3:5, 3:4 oder 4:5 stehen. Würde der erste Ton z.B. der Ton *c* sein, so ergäben sich bereits 5 Tonigkeiten des diatonischen Tonsets, nämlich die Tonigkeiten *c* (1:1), *g* (2:3), *a* (3:5), *f* (3:4), *e* (3:4). Die fehlenden zwei Tonigkeiten *d* und *h* könnten dann noch als Quinte und Terz der zur Tonigkeit *c* quintverwandten Tonigkeit *g* hergeleitet werden. Die genauen Grundfrequenzen der Tonigkeiten des diatonischen Tonsets hängen sowohl von der gewählten Bezugstonart als auch der gewählten *Stimmung* ab (Abschnitt 2.6.3). Die *Benennung* der Tonigkeiten eines diatonischen Tonsets ergibt sich aus deren pythagoreischer Herleitung. Diese wird in Abschnitt 2.6.3 behandelt.

Identifikation von diatonischen Tonsets

Da das Tonset gegenüber der Tonleiter in der musiktheoretischen Literatur eine untergeordnete Rolle spielt, gibt es noch kein System zur Identifikation und Benennung diatonischer Tonsets. Die Autoren verwenden deshalb folgende Konventionen: Die *Benennung* eines diatonischen Tonsets erfolgt über den *Namen der daraus erzeugbaren Durtonleiter*. Wenn vom diatonischen Tonset C-Dur geredet wird, dann sind die Tonigkeiten *c*, *d*, *e*, *f*, *g*, *a* und *h* gemeint. Aus diesen Tönen können natürlich auch die anderen Modi *a*-Moll/Äolisch, *d*-Dorisch, *e*-Phrygisch, *f*-Lydisch, *g*-Mixolydisch und *h*-Lokrisch erzeugt werden. Die *numerische Identifikation* eines diatonischen Tonsets erfolgt über einen Tonsetindex n_k .²⁶ Die Variable n_k repräsentiert dabei in Anlehnung an [47] die Anzahl der gegenüber dem diatonischen Tonset C-Dur erhöhten oder erniedrigten Töne und damit die Anzahl der Vorzeichen. Positive n_k repräsentieren eine Anzahl von gegenüber C-Dur erhöhten Tönen bzw. Kreuzen (\sharp) und negative n_k eine Anzahl von gegenüber C-Dur erniedrigten Tönen bzw. Bs (*b*). Das diatonische Tonset F-Dur besitzt z.B. ein *b* und hat damit einen Tonartindex von $n_k = -1$. Das diatonische Tonset G-Dur besitzt ein \sharp und hat damit einen Tonartindex von $n_k = +1$. Weitere Zuordnungen sind in Tabelle 2.3 zu finden.

²⁶ Das *k* steht für den englischen Term *key*.

<i>Tonset</i>	...	<i>As</i>	<i>Es</i>	<i>B</i>	<i>F</i>	<i>C</i>	<i>G</i>	<i>D</i>	<i>A</i>	<i>E</i>	...
n_k	...	-4	-3	-2	-1	0	+1	+2	+3	+4	...

Tabelle 2.3 – Zuordnung von Tonsetindizes n_k zu diatonischen Tonsets K

2.6.2 Das chromatische Tonset

Das chromatische Tonset ist eine Obermenge des diatonischen Tonsets. Es besteht aus zwölf Tonigkeiten, wobei der Tonhöhe entsprechend benachbarte Tonigkeiten einen Abstand von einem Halbtonschritt bzw. einer kleinen Sekunde haben. Das chromatische Tonset kann hergeleitet werden, indem man z.B. die diatonischen Tonsets ausgehend von quintbenachbarten Tonigkeiten bestimmt und zusammenführt. Die genauen Grundfrequenzen der Töne sind auch hier von der gewählten Stimmung und der gewählten Bezugstonart abhängig.

2.6.3 Stimmung

Um die genauen Frequenzen der Tonigkeiten von chromatischen oder diatonischen Tonsets zu bestimmen, ist es erforderlich, die genaue Stimmung festzulegen. Jedes Stimmungssystem stellt im Grunde genommen einen Kompromiss dar: Entweder stimmt man die Töne so, dass sich sehr reine Intervalle²⁷ ergeben, oder man stimmt die Töne so, dass man in möglichst vielen Tonarten spielen kann. In der Musikgeschichte haben sich verschiedene Stimmungssysteme herausgebildet, wobei hier nur vier kurz betrachtet werden sollen.

- *Die pythagoreische Stimmung*: Die zwölf chromatischen Tonigkeiten ergeben sich, indem man von einer Referenztonigkeit sechs reine Quinten in beide Richtungen stimmt²⁸. Das Problem dieser Stimmung besteht darin, dass sich Anfang und Ende der so hergeleiteten Tonigkeitsreihe um ca. $\frac{1}{4}$ Halbton, das sogenannte *pythagoreische Komma*²⁹, unterscheiden und sich somit nicht zu einem Kreis schließen [112, S. 88-90]. Dadurch sind Akkorde, welche Tonigkeiten aus beiden Enden der Quintreihe

²⁷ Unter reinen Intervallen versteht man Intervalle, die in einfachen Frequenzverhältnissen gestimmt sind, z.B. 2:3, 3:4 etc.

²⁸ Z.B. kann man von der Tonigkeit *c* durch Stimmen in reinen Quinten aufsteigend die Tonigkeiten *g, d, a, e, h* und *fis* und durch Stimmen in Quinten absteigend die Tonigkeiten *f, b, es, as, des, ges* bestimmen.

²⁹ Als pythagoreisches Komma wird das Frequenzverhältnis $(3:2)^{12} : (2:1)^7 = 1,014$ bezeichnet, was in etwa einem viertel Halbton entspricht [112, S. 88-90].

verwenden, sehr dissonant. Ein weiteres Problem der pythagoreischen Stimmung ist die Unreinheit der für Dur- und Molldreiklänge so wichtigen Terz.³⁰

- *Die reine Stimmung*: Die reine Stimmung versucht das Problem der unreinen Terz der pythagoreischen Stimmung zu lösen. Die chromatischen Tonigkeiten ergeben sich dadurch, dass von einer gegebenen Tonigkeit zunächst in reinen Quinten gestimmt wird und anschließend zu jeder der sich so ergebenden Tonigkeiten die reine große Terz gebildet wird^{31,32}. Der Vorteil der reinen Stimmung besteht in der Intervallreinheit der Hauptdreiklänge der gewählten Tonart (Abschnitt 5.3.1). Jedoch geht diese Intervallreinheit immer mehr verloren, je weiter sich das Musikstück von der Ausgangstonart entfernt. Grenzenloses Modulieren in andere Tonarten ist deshalb mit der reinen Stimmung nicht möglich [112, S. 88-90].
- *Die mitteltönige Stimmung*: Wie die reine Stimmung soll die mitteltönige Stimmung das Problem der unreinen Terz in der pythagoreischen Stimmung zu lösen. Das Problem hierbei besteht darin, dass sich die aus vier Quinten ergebende große Terz um einen als *syntonisches Komma* bezeichneten Faktor $\frac{81}{80}$ von der reinen Terz unterscheidet und damit etwas zu groß ist. Die mitteltönige Stimmung löst dieses Problem, indem sie die Intervalle in der Quintreihe um ein Viertel des syntonischen Kommas verkleinert, so dass vier aufeinander folgende Quinten eine reine Terz ergeben.³³ Dadurch wird der spielbare Tonartenbereich vergrößert. Jedoch schließt sich auch die mitteltönige Quintreihe nach zwölf Quinten nicht zu einem Kreis, so dass bestimmte Tonarten nicht gespielt werden können.

³⁰ Ein Beispiel: Die Tonigkeiten der reinen Terz *c-e* haben das Frequenzverhältnis $\frac{5}{4} = 1,25$. In der pythagoreischen Stimmung aber ergibt sich die Terz *c-e* über vier Quinten (*c-g-d-a-e*) und hat das Frequenzverhältnis $\frac{3^4}{2} * 1/4 = \frac{81}{64} = 1,266$. Das Verhältnis aus pythagoreischer Terz und reiner Terz $\frac{81}{64} : \frac{5}{4} = \frac{81}{80}$ wird als syntonisches Komma bezeichnet.

³¹ Die chromatischen Tonigkeiten bezogen auf die Tonigkeit C ergeben sich damit folgendermaßen: In Quinttrichtung: *b-f-c-g-d-a*. Jeweils in Terzrichtung: *d-a-e-h-fis-cis-gis*. Hier ist bereits zu sehen, dass die Tonigkeit *d* einmal als Quinte der Tonigkeit *g* und einmal als große Terz der Tonigkeit *b* vorhanden ist. Das Frequenzverhältnis zwischen diesen beiden Tonigkeiten wird als *syntonisches Komma* bezeichnet [112, S. 89]. (Siehe auch Abschnitt 2.6.3 und Abschnitt 3.8).

³² Die Idee der reinen Stimmung wird im Euler-Netz sehr schön veranschaulicht (Abschnitt 3.4, Abbildung 3.4).

³³ Ein Beispiel: Die reine große Terz *c-e* besitzt jedoch das Frequenzverhältnis $\frac{5}{4}$. Die pythagoreische große Terz *c-e* lässt sich aus den Quinten *c, g, d, a* und *e* hergeleitet und besitzt das Frequenzverhältnis $\frac{3^4}{2} * 1/4 = \frac{81}{64}$ und ist damit um einen Faktor $\frac{81}{64} : \frac{5}{4} = \frac{81}{80}$, dem syntonischen Komma, zu groß. Verringert man nun jede Quinte der Quintreihe *c, e, g, d, a* um ein Viertel des syntonischen Kommas ($\frac{81}{80}^{\frac{1}{4}}$), so ergeben diese exakt eine reine Terz: $(\frac{3}{2} / \frac{81}{80}^{\frac{1}{4}})^4 * 1/4 = 1.25$.

- Die temperierte Stimmung³⁴ verteilt die zwölf chromatischen Tonigkeiten in logarithmisch gleichmäßigen Abständen auf den Frequenzbereich einer Oktave. Zwei chromatisch benachbarte Tonigkeiten haben dadurch das Frequenzverhältnis $\frac{f_1}{f_0} = 2^{\frac{1}{12}}$. Das Frequenzverhältnis der temperierten Quinte ist dadurch $2^{7/12} = 1,498$, was leicht von der reinen Quinte ($3/2 = 1,5$) abweicht. Der Vorteil der temperierten Stimmung besteht darin, dass sich die pythagoreische Quintreihe am Ende schließt. Dadurch sind die Frequenzverhältnisse der Akkorde aller Tonarten identisch, was es möglich macht, auch in entfernte Tonarten zu wechseln (zu modulieren).

2.6.4 Kirchentonleitern und weitere Skalen

Aus den sieben Tonigkeiten des diatonischen Tonsets ergeben sich je nach Wahl des Anfangs- bzw. Endtones die in Tabelle 2.4 aufgelisteten Modi bzw. Kirchentonarten.³⁵

Modus	Ionisch	Dorisch	Phrygisch	Lydisch	Mixolydisch	Äolisch	Lokrisc
Anfangston	C	d	e	F	G	a	h

Tabelle 2.4 – Die sieben Modi aufbauend auf dem diatonischen Tonset C-Dur (nach [62, S. 71])

Die Kirchentonarten spielten im Mittelalter eine große Rolle, wobei dies jedoch nur für die Modi Dorisch, Phrygisch, Lydisch und Mixolydisch gilt. Die Modi Ionisch und Äolisch wurden erst zu Beginn der Neuzeit ergänzt [84, S. 9]. Interessant hierbei ist, dass gerade diese beiden Kirchentonarten, die anfänglich so geringe Bedeutung hatten, im 17. Jahrhundert als Dur- und Molltonleiter die ursprünglichen vier Kirchentonleitern verdrängten. Im 20. Jahrhundert bekamen jedoch alle Kirchentonleitern als Jazz-Skalen im Bereich der „kreativen Improvisation“ [62, S. 70] oder auch als „modale Bereicherung der tonalen Qualitäten“ [84, S. 9] wieder neue Bedeutung. In diesem Zusammenhang wurden die sechs Kirchentonarten um den Modus „Lokrisc“ erweitert. Eine Einordnung der Kirchentonarten in das Symmetriemodell ist in Abschnitt 5.5.1 zu finden.

Neben den diatonischen Kirchentonleitern gibt es verschiedene Modifikationen der Diatonik durch Veränderung (Erhöhung oder Erniedrigung) einzelner Tonigkeiten um einen Halbton. Die populärste Abwandlung der Diatonik ist die harmonische Molltonleiter, bei

³⁴ Die temperierte Stimmung wird auch als gleichtemperierte, gleichstufige, gleichschwebende oder zwölftemperierte Stimmung bezeichnet.

³⁵ Neben den authentischen Kirchentonarten gibt es noch die plagalen Kirchentonarten [112, S. 91], [84, S. 9], welche hier nicht näher betrachtet werden.

welcher der siebente Ton der Tonleiter um einen Halbton erhöht wurde, so dass eine Leittonwirkung [84, S. 67] entsteht (siehe auch Abschnitt 5.5.2). Ein Überblick über weitere Skalen und deren Eigenschaften kann z.B. in [69, Chapter 4] gefunden werden.

2.6.5 Leitereigene und leiterfremde Klänge

Klänge, die aus dem Material eines gegebenen Tonsets erzeugt werden können, werden als *leitereigene Klänge* bezeichnet, Klänge, die auch Töne oder Tonigkeiten anderer Tonsets verwenden, als *leiterfremde Klänge*. In der Wahrnehmung werden leitereigene Akkorde viel stärker als zusammengehörend empfunden als leiterfremde Klänge (siehe Abschnitt 2.7.2). Die Verwendung leitereigener Akkorde stellt den Standardfall dar, leiterfremde Akkorde dagegen werden in Form von Zwischendominanten, Ausweichungen usw. eingesetzt, um den harmonischen Kontrast eines Musikstückes zu erhöhen.

2.7 Kontextbezogene Wahrnehmung von Tonalität

Bisher wurde die Wahrnehmung von Einzeltönen, Intervallen und Akkorden für den Fall betrachtet, dass diese isoliert erklingen. In der Praxis jedoch treten musikalische Ereignisse selten allein auf, sondern meist in einem sogenannten *Kontext*, d.h. im Zusammenhang mit anderen Klängen. Das auditorische System des Menschen versucht, alle wahrgenommenen Ereignisse zueinander in Beziehung zu setzen und in eine schlüssige musikalische Gestalt zu integrieren, ein Prozess, der sich stark auf die qualitative Wahrnehmung eines Musikstückes auswirkt. Gegenstand dieses Abschnittes sind deshalb verschiedene Grundlagenexperimente, welche die Entstehung des Kontexts sowie dessen Einfluss auf die Wahrnehmung von Einzeltönen und Akkorden demonstrieren. Auf Basis der hier gelegten Grundlagen wird es später möglich sein, geometrische Tonräume in einen wahrnehmungspsychologischen Bezug zu setzen und Maße für die harmonieorientierte Ähnlichkeit bzw. Nähe von tonalen Ereignissen abzuleiten.

2.7.1 Tonale Hierarchien und tonale Stabilität

1982 veröffentlichte Carol L. Krumhansl [80] die Ergebnisse eines Experiments, in welchem sie die Wahrnehmung von einzelnen Tönen in einem gegebenen Kontext demonstrierte. In diesem Experiment hörten musikalisch erfahrene Probanden zunächst eine Akkordkadenz oder eine Tonleiter, welche von einem sogenannten Testton (Probe-Tone) gefolgt wurde. Die Aufgabe der Testpersonen bestand darin zu bewerten, wie gut der am Ende folgende Ton zu der vorher gespielten Ton- oder Akkordfolge passt. Dies

wurde für alle zwölf chromatischen Tonigkeiten durchgeführt. Ergebnisse des Experimentes sind die in Abbildung 2.5 gezeigten Testtonprofile, welche die durchschnittlichen Bewertungen aller zwölf chromatischen Tonigkeiten wiedergeben und zwar a) für einen C-Dur Kontext und b) für einen c-Moll Kontext. Das interessanteste Ergebnis dieses Experimentes besteht in der Ähnlichkeit der Bewertungen unterschiedlicher Probanden. Wie es scheint, baut das auditorische System des Menschen eine sogenannte *tonale Hierarchie* auf, welche bestimmten Gesetzen folgt. Die Position von Tonigkeiten in dieser Hierarchie ist ein Maß für die *tonale Stabilität*. Wie in Abbildung 2.5 gesehen werden kann, ist der Grundton der Kontexttonart am stabilsten. Er wird deshalb auch als *tonales Zentrum* bezeichnet. Alle anderen Töne sind auf diesen Ton bezogen. Weiterhin sehr stabil sind die Terz und die Quinte des Grundtones. Zusammen mit dem Grundton bilden diese Tonigkeiten auch den sogenannten *Tonikadreiklang* (Abschnitt 5.3.1). In der musikalischen Praxis findet sich die tonale Stabilität sowohl in der *Auftrittswahrscheinlichkeit* von Tönen als auch in deren zeitlicher Position wieder: Anhand von Ergebnissen von H. Budge [18] zeigte Krumhansl weiterhin [82], dass die ermittelten Profile in einem starken Maß mit den Tonauftrittswahrscheinlichkeiten in Musikstücken des 18. und 19. Jahrhunderts korrelieren. Anfang und Ende vieler Musikstücke wird in den meisten Fällen durch stabile Tonigkeiten gebildet, das ist die Tonika. Was die *Stabilität von Akkorden* betrifft, fand Krumhansl in weiteren Experimenten heraus [82, S. 168ff], dass die wahrgenommene Stabilität von Akkorden der Stabilität von Einzeltönen sehr ähnlich ist und durch Bilden der Mittelwerte der Stabilitäten der Akkordtöne berechnet werden kann (S. 177). Weiterhin ergab sich, dass Mollakkorde in der Regel weniger stabil als Durakkorde sind und die harmonische Funktion (Abschnitt 5.3 von Mollakkorden weniger stark ausgeprägt ist.³⁶

2.7.2 Wahrgenommene Nähe von Tönen und Akkorden

In einem weiteren Experiment untersuchte Krumhansl die kontextabhängige, wahrgenommene Nähe zweier Töne [81]. Der Kontext wurde zunächst durch Spielen eines Dur-Akkordes bzw. einer auf- und absteigenden Tonleiter etabliert. Anschließend wurden zwei Töne sequentiell präsentiert. Die Probanden hatten zu bewerten, wie gut der zweite Ton des Paares zum ersten Ton passt. Das Experiment wurde für alle möglichen Tonpaare

³⁶ Dies entspricht Terhardts Theorie der virtuellen Tonhöhe (Abschnitt 2.3.2). Durakkorde stimmen stark mit dem Obertonspektrum überein und ermöglichen deshalb eine eindeutige Bestimmung des virtuellen Grundtones, was bei Mollakkorden nicht der Fall ist.

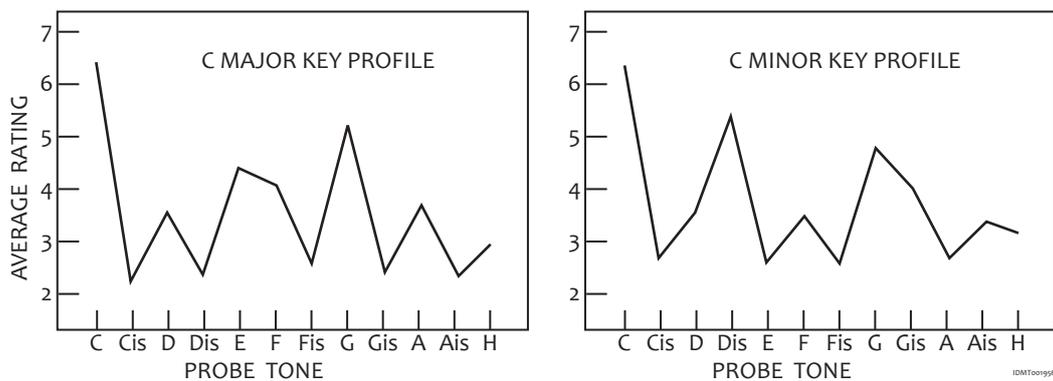


Abbildung 2.5 – Carol L. Krumhansls Testtonprofile für (a) C-Dur und (b) c-Moll [80]

durchgeführt. Die folgende Multidimensional Scaling Analyse (MDS)³⁷ der gewonnenen Daten ergab den in Abbildung 2.6 dargestellten Kegel, welcher die zwölf chromatischen Tonigkeiten auf drei Ebenen organisiert: Die erste Ebene (Spitze des Kegels) besteht aus den Tonigkeiten des Tonikadreiklages, die zweite Ebene wird durch die verbleibenden leitereigenen diatonischen Tonigkeiten und die dritte Ebene durch die restlichen chromatischen Tonigkeiten gebildet. Wie es scheint, werden die zwölf Tonigkeiten auf eine akkordische, eine diatonische und eine chromatische Ebene aufgeteilt. Die Tonhöhe der Töne findet sich auf der Winkeldimension des Modells wieder. Insgesamt hebt das Modell deutlich hervor, dass die wahrgenommene Nähe zweier Töne durch den Kontext beeinflusst wird. So werden im Kontext von C-Dur die Töne *fis* und *ais* (große Terz) als deutlich weiter voneinander entfernt wahrgenommen als die Töne *c* und *e* (ebenfalls große Terz). Dieses und das zuvor beschriebene Experiment inspirierten Fred Lerdahls Tonraummodell (Abschnitt 3.7) stark.

Um auch die wahrgenommene Nähe von Akkorden in einem gegebenen Kontext zu bestimmen, führte Krumhansl ein dem vorherigen Test ähnliches Experiment durch [82, S. 188ff]. Dazu wurde zunächst in Form von Tonleitern und Akkordkadenzen eine der Tonarten *C*, *G*, *A*, *H* oder *Fis*-Dur etabliert. Es folgte ein Akkordpaar aus den Tonarten C-Dur oder *Fis*-Dur. Aufgabe der Probanden war wieder zu bewerten, wie gut der zweite Akkord zum ersten passt. Die Bewertungsergebnisse wurden wieder mit Hilfe von MDS analysiert. Ergebnis des Experimentes war, dass zwei Akkorde als näher zueinander wahrgenommen wurden, wenn sie zur etablierten Tonart gehörten, d.h. leitereigen waren. Tonartfremde Akkorde dagegen wurden als weniger zusammengehörend empfunden.

³⁷ Multidimensional Scaling ist ein statistisches Verfahren, das die Ähnlichkeiten zwischen Objekten in einem niedrigdimensionalen geometrischen Raummodell darstellt.

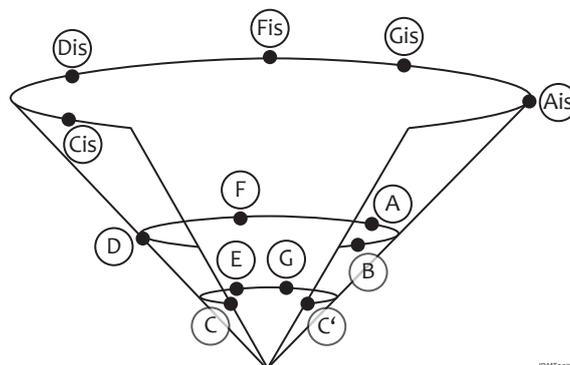


Abbildung 2.6 – Wahrnehmere Nähe von Tönen im C-Dur Kontext (nach [81])

2.7.3 Wahrnehmung von Tonarten

In welcher Form kann die Wahrnehmung von Tonarten beschrieben werden? Eine Möglichkeit hierfür sind die in Abschnitt 2.7.1 eingeführten Testtonprofile. Diese zeigen, wie stark eine Tonart etabliert ist. Um herauszufinden, wie die Etablierung einer Tonart in der Wahrnehmung erfolgt, erzeugten Carol L. Krumhansl und Edward J. Kessler [80] aus neun Akkorden bestehende Sequenzen. Jedoch wurde den Testpersonen zunächst nicht die ganze Akkordsequenz präsentiert, sondern nur ein einzelner Akkord. Diesem folgte einer von zwölf chromatischen Testtönen, welcher von den Testpersonen wie im Experiment aus Abschnitt 2.7.1 bewertet werden musste. Nach Erhebung des gesamten Testtonprofils für den ersten Akkord der Sequenz wurde das Experiment wiederholt, jedoch mit den ersten beiden, drei usw. Akkorden der Sequenz. Ergebnis dieser Prozedur war das Vorliegen von Testtonprofilen für jeden Zeitpunkt der Akkordfolge. Die Auswertung der Testtonprofile ergab eine hohe Übereinstimmung der Ergebnisse unterschiedlicher Probanden. Es stellte sich heraus, dass sich der Sinn für eine Tonart nach und nach aufbaut: Je mehr Akkorde einer Tonart erklingen, desto klarer bildet sich die tonale Hierarchie, wie in Abschnitt 2.7.1 beschrieben, heraus und desto stärker korreliert das zu einem Zeitpunkt gemessene Testtonprofil mit den in Abschnitt 2.7.1 beschriebenen Testtonprofilen. Weiterhin ist interessant, dass bereits das Akkordpaar *IV-V* (Subdominante-Dominante) eine klare tonale Hierarchie etabliert. Das Erklingen der *I* (Tonika) ist dazu nicht erforderlich [80, Abbildung 8.1]. Bei Modulationen kommt es beim Erklingen von Akkorden der neuen Tonart nicht sofort zu einer abrupten Umstellung der tonalen Hierarchie. Vielmehr wird die alte Hierarchie erst nach und nach abgebaut. Gleichzeitig kommt es zum Aufbau der neuen Tonart. Gerade bei Modulationen in weiter entfernte Tonarten gibt es hier Übergangsbereiche, innerhalb derer weder die eine noch die andere Tonart vollständig etabliert ist.

Sehr wichtig für den perzeptuellen Vergleich von Musikstücken ist auch die Möglichkeit, die Wahrnehmbarkeit von Tonartwechseln bestimmen zu können. Ein Komponist kann auf der einen Seite ein Musikstück so komponieren, dass Modulationen kaum wahrnehmbar sind. Auf der anderen Seite ist es möglich, dass es bei Modulationen zu einem deutlichen „tonalen Ruck“ kommt [83, S. 63]. Um derartige semantische Eigenschaften des Musikstückes analysieren zu können, ist es erforderlich, die *wahrgenommene Ähnlichkeit* von Tonarten zu bestimmen. Diese ergibt sich zum einen aus der Anzahl der gemeinsamen Töne, was zu einer dem Quintenzirkel entsprechenden Anordnung von Tonarten – wie in Abschnitt 3.1 beschrieben – führt. Krumhansls tonale Hierarchien (Abschnitt 2.7.1) zeigen jedoch, dass es ein weiteres Merkmal für die Ähnlichkeit von Tonarten gibt: die Stabilität der die Tonart bildenden Tonigkeiten.³⁸ Krumhansl und Kessler berechneten deshalb die Korrelation zwischen den Testtonprofilen aller Tonarten und leiteten so ein Ähnlichkeitsmaß für jedes Paar von Dur- und Molltonarten her. Die MDS-Analyse der resultierenden Daten ergab einen vierdimensionalen Torus, welcher in Abbildung 2.7 dargestellt ist. Die Abstände der Tonarten auf der Oberfläche des Torus stellen die wahrgenommenen Distanzen dar. Wie in Abbildung 2.7 zu sehen, finden sich einige bereits aus der Musiktheorie bekannte Zusammenhänge wieder, darunter die Benachbarung von parallelen und gleichnamigen Dur- und Molltonarten oder auch die Quintverwandtschaft von Tonarten. Vom Aufbau ist Krumhansls 4D MDS-Lösung sehr ähnlich dem Vial/Weber/Schönberg-Netz, welches in Abschnitt 3.2 ausführlich behandelt wird.

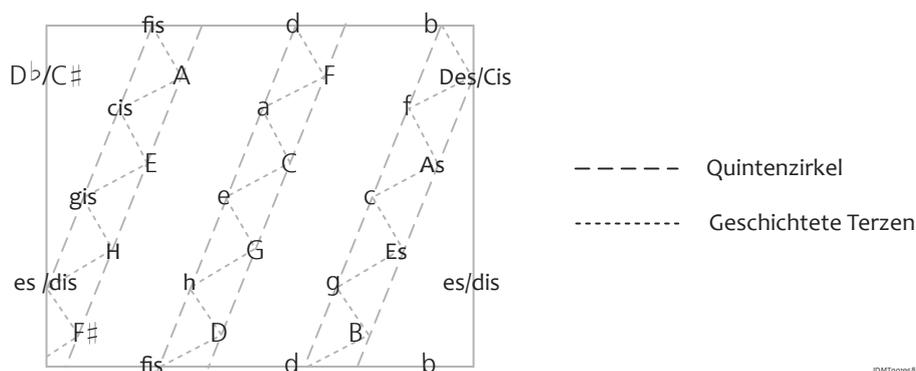


Abbildung 2.7 – Zweidimensionale Darstellung der vierdimensionalen MDS-Lösung zur Ähnlichkeit von Tonarten (nach [82])

³⁸ Die Tonarten C-Dur und c-Moll besitzen z.B. nur vier gemeinsame Töne und sind im Quintenzirkel drei Tonarten voneinander entfernt. Jedoch besteht eine hohe Korrelation zwischen den Testtonprofilen beider Tonarten, was eine mögliche Erklärung für das häufige Auftreten von Modulationen in die gleichnamige Moll- bzw. Durtonart ist.

2.7.4 Perzeptuelle Asymmetrien

Ein Problem der bis hierher dargestellten geometrischen Tonalitätsmodelle ist die mangelnde Fähigkeit, perzeptuelle Asymmetrien darstellen zu können. Als perzeptuelle Asymmetrie bezeichnet man das Phänomen, dass bestimmte Eigenschaften der Wahrnehmung musikalischer Ereignisse von der Reihenfolge ihres Auftretens abhängt. Beispiele hierfür sind z.B. Experimente, die zeigten, dass Modulationen in Tonarten im Quintenzirkel aufwärts als näher empfunden wurden als Abwärtsmodulationen [82, S. 49]. Ein anderes Beispiel sind der Leitton und der Grundton: Eine Schreitbewegung vom Leitton zum Grundton (h nach c) wird als wesentlich kürzer empfunden, als ein Sprung in die umgekehrte Richtung. In geometrischen Modellen wie z.B. dem in Abbildung 2.6 wird dies aber nicht berücksichtigt. Eine mögliche Lösung dieses Problems besteht in der Einordnung von Kräftenmodellen in geometrische Tonräume. Solche Kräftenmodelle könnten z.B. dem Grundton einer Tonart eine tonale Anziehungskraft zuordnen, welche wiederum auf andere Töne wie z.B. dem Leitton wirkt. Erste Ansätze in dieser Richtung wurden z.B. von Fred Lerdahl (Tonal Pitch Space) entwickelt (Abschnitt 3.7).

2.7.5 Funktionstheoretische Aspekte

Die musiktheoretische Erfassung kontextbezogener Aspekte der Tonalitätswahrnehmung erfolgte mit Hugo Riemanns *Funktionstheorie* [134], welche an dieser Stelle beschrieben werden müsste. Da sich aber viele Eigenschaften dieser Theorie im *Symmetriemodell*, welches in dieser Arbeit entwickelt wird, sehr einfach darstellen lassen, wird hier auf Kapitel 4 verwiesen.

Kapitel 3

Stand der Wissenschaften

Tonraummodelle

In den nachfolgenden Absätzen geht es um den Stand der Technik im Bereich von geometrischen Tonraummodellen. Wir werden mit der Darstellung hauptsächlich aus der Musiktheorie stammender Tonraumsysteme beginnen. Neben der Diskussion der hinter den jeweiligen Tonräumen stehenden Konzepte soll auch ein Bezug zur Wahrnehmungspsychologie hergestellt werden. Während die Konzepte für die Extraktion von Highlevel-Features wie Tonarten, Akkorden oder Melodien wichtig sind, ist der wahrnehmungspsychologische Bezug der diskutierten Systeme für den perzeptuellen Vergleich von Musiksignalen von Bedeutung. Nur wenn ein System die tatsächlich wahrgenommenen Ähnlichkeiten widerspiegelt, kann es gute Ergebnisse beim Vergleich von Musikstücken liefern. Übergeordnetes Ziel des gesamten Kapitels ist die Beantwortung folgender Fragen:

1. Inwiefern ist das jeweilige Tonraummodell für die Analyse von realem Audio (vgl. Kapitel 7) geeignet? Ist eine Anwendung für die Analyse obertonreicher, rausch- bzw. störsignalbehafteter Audiosignale möglich? Wie können derartige Signale in dem jeweiligen Tonraum abgebildet werden?
2. Inwiefern ist das jeweilige Tonraummodell für den Ähnlichkeitsvergleich von Musiksignalen geeignet? Welche technischen Beschreibungsparameter können aus dem jeweiligen Tonraum extrahiert werden?
3. Inwiefern ist es möglich, mit einem jeweiligen Tonraum abstrakte musikalische Merkmale wie Tonarten, Akkorde, Noten etc. zu extrahieren? Ein besonderer Schwerpunkt wird hierbei auf die Anwendungsgebiete Tonart- und Akkorderkennung gelegt.

4. Welche Möglichkeiten bietet ein Tonraummodell zur Synthese musikalischer Audiosignale, d.h. für die Entwicklung tonraumbasierter Musikinstrumente, wie in Kapitel 6 beschrieben?
5. Inwiefern ist ein Tonraummodell zur Vermittlung musiktheoretischer Grundkonzepte geeignet? Die Diskussion dieses Kriteriums zielt auf einen Einsatz in der Musikpädagogik ab.

3.1 Heinichens, Matthesons und Kellners musikalische Kreise

Geometrische Analogien werden schon lange verwendet, um musikalische Sachverhalte zu veranschaulichen. Es ist anzunehmen, dass eines der ersten und schon früh verbreiteten geometrischen Tonraummodelle der Quintenzirkel war. Dies geht z.B. aus [65, S. 838] hervor, wo David Heinichen 1728 den Quintenzirkel als „die bekandte Arth des Kircheri, alle Modos per Ovartas & Qvintas zu circuliren“ als „eine der unvollkommens-ten“ bezeichnete und einen „Musicalischen Circul“ einführte (Abbildung 3.1a). Ziel Heinichens war es, die „natürliche Ordnung, Verwandtschaft und Ausschweifung“ aller Tonarten („Modorum Musicorum“) darzustellen (S. 837). Der Kreis sollte es sowohl Instrumentalisten als auch Komponisten ermöglichen, den Überblick auch bei Modulation in entfernte Tonarten zu behalten. Eine Schwäche des Quintenzirkels bestand für Heinichen darin, dass dieser nur die Verwandtschaft von Tonarten gleichen Tongeschlechts, nicht jedoch die Verwandtschaft von Tonarten unterschiedlichen Tongeschlechts abbildet. Heinichen stellt deshalb sowohl die 12 Dur- als auch die 12 Molltonarten als zwei ineinander verwobene Quintenzirkel in einem einzigen geometrischen System dar, wobei die Verwandtschaft zwischen Dur- und paralleler Molltonart durch geometrische Benachbarung hervorgehoben wird. Dadurch entsteht jedoch das Problem, dass die neben der Parallelverwandtschaft von Tonarten so wichtige Quintverwandtschaft nicht mehr geometrisch dargestellt werden kann. Johann Mattheson veröffentlicht deshalb 1735 einen „verbesserten Musicalischen Kreis“ (Abbildung 3.1b), der sich von Heinichens Kreis dahingehend unterscheidet, dass sich jeweils ein Paar quintverwandter Tonarten mit einem Paar terzverwandter Tonarten abwechseln. Jedoch kann Heinichen damit nur die Quintverwandtschaft jedes zweiten Tonartenpaares repräsentieren¹, ein Problem, das 1737 von David Kellner gelöst wird: Kellners Tonraum repräsentiert Dur- und Molltonarten in zwei getrennten Quintenzirkeln, wobei der Mollquintenzirkel im Inneren des Durquintenzir-

¹ So wird z.B. die Quintverwandtschaft zwischen C-Dur und G-Dur durch geometrische Benachbarung veranschaulicht, nicht jedoch die Quintverwandtschaft zwischen F-Dur und C-Dur.

kels dargestellt wird. Durch diesen Schritt gelingt es sowohl die Quintverwandtschaft als auch die Parallelverwandtschaft von Tonarten durch geometrische Benachbarung abzubilden.²



Abbildung 3.1 – Erste Ansätze zur geometrischen Repräsentation von Modulationen

Bewertung:

Heinrichs- und Matthesons Kreise zeigen die Anfänge geometrischer Tonraummodelle, geben jedoch die wahrgenommenen Ähnlichkeiten von Tonarten (siehe Abschnitt 2.7.3) nur teilweise wieder. Kellners System spiegelt zwar die Quint- und die Parallelverwandtschaft von Tonarten gut wider, nicht jedoch die ebenfalls wichtige Variantverwandtschaft. Ein Vergleich des tonalen Verlaufs von Musikstücken auf Basis der in den zuvor diskutierten Kreisen zugrundeliegenden geometrischen Verhältnisse wird deshalb nur zu unbefriedigenden Ergebnissen führen.

3.2 Das Vial/Weber/Schönberg-Netz

Aufgrund der zuvor beschriebenen Nachteile von Kellners musikalischem Kreis, entwickelte F. G. Vial ein zweidimensionales Tonartennetz ([164], vgl. [101, S. 42] und [50]), das jedoch als Gottfried Webers [166] „Tabelle der Tonartenverwandtschaften“ (Abbildung 3.2c) oder auch „Schönberg-Netz“ [144, S. 20] bekannt ist. Wie in Abbildung 3.2

² In Benachbarung zur Tonart C-Dur befinden sich die quintverwandten Tonarten F-Dur und G-Dur als auch die parallele Molltonart a-Moll.

zu sehen ist, werden Tonarten als Knotenpunkte eines Netzes dargestellt. Oberhalb und unterhalb einer jeden Tonart befinden sich deren quintverwandte Tonarten, links und rechts die Parallel- und die gleichnamige Tonart entgegengesetzten Tongeschlechts. Weber unterscheidet dabei zwischen verwandten Tonarten 1. Grades, das sind die im Netz horizontal und vertikal direkt benachbarten Tonarten, und verwandten Tonarten 2. Grades, das sind die direkten Nachbarn der Verwandten ersten Grades.

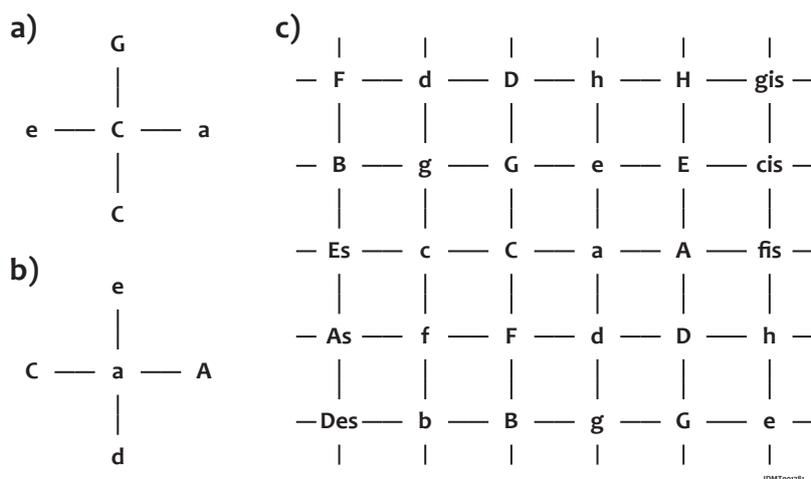


Abbildung 3.2 – Anordnung von Tonarten im Vial/Weber/Schönberg-Netz [166]

Bewertung:

Das Vial/Weber/Schönbergnetz bildet die drei wichtigsten Grundverwandtschaften von Tonarten ab (Quint-, Parallel- und Variantverwandtschaft) und kommt vor allem den von Krumhansl und Kessler empirisch ermittelten Ähnlichkeiten von Moll- und Durtonarten nahe (Abschnitt 2.7.3). Ein Vergleich von Musikstücken auf Basis der in diesem Raum eingezeichneten Pfade kann deshalb zu guten Ergebnissen führen. Eine wünschenswerte Erweiterung des Netzes wäre die Abbildung von Kirchentonarten. Diese verloren während der dur-moll-tonalen Epoche zwischen 1600 und 1900 ihre Bedeutung, gewannen aber als „modale Bereicherung der tonalen Qualitäten“ im 20. Jahrhundert wieder größere Aufmerksamkeit [84, S. 9] (siehe auch Abschnitt 2.6.4). Um reale Audiosignale mit Hilfe des Vial/Weber/Schönbergnetzes zu analysieren, ist eine vorgeschaltete dynamische Tonarterkennung notwendig, welche den Tonartenverlauf eines Musikstückes über die Zeit analysiert und in das Netz einzeichnet. Schwierig ist hierbei jedoch, dass die Tonart in vielen Fällen nicht eindeutig bestimmt werden kann. Eine Erweiterung des Vial/Weber/Schönbergnetzes für mehrdeutige tonale Situationen ist deshalb notwendig. Für die Synthese von Musiksignalen stellt das Netz eine mögliche Grundlage für Benutzerschnittstellen zur Eingabe und zum Wechsel von Tonarten dar.

3.3 Der Eulerraum

Die Intention der bisher analysierten Tonräume war die Abbildung der Ähnlichkeiten von Tonarten. Eine weitere Anwendung geometrischer Tonräume ist die Herleitung der im abendländischen Bereich sehr verbreiteten reinen Stimmung. Ausgehend von den ersten Harmonischen des Obertonspektrums eines periodischen Signals und einer gewählten Frequenz f_0 schlug der Mathematiker Leonard Euler 1739 einen Tonraum vor [41], welcher die Oktavobertöne $\dots, 2f_0, 4f_0, 8f_0, \dots$ auf der ersten, die Quintobertöne $\dots, 3f_0, 9f_0, 27f_0, \dots$ auf der zweiten und die Großterzobertöne $\dots, 5f_0, 25f_0, \dots$ auf der dritten Achse eines dreidimensionalen Koordinatensystems anordnete (Abbildung 3.3). Die einer reinen Stimmung zugrundeliegenden Tonfrequenzen (Abschnitt 2.6.3) konnten damit als diskrete Schnittpunkte der Oktav-, Quint- und Terzlinien des Eulerraumes hergeleitet und mit der Formel $f = f_0 * 2^p * 3^s * 5^r$ ermittelt werden [109]. Beim Eulerraum muss beachtet werden, dass zwei auf der Terz- oder Quintachse benachbarte Punkte nicht mit der Terz ($f_1/f_0 = 3/2$) oder der Quinte ($f_1/f_0 = 5/4$) verwechselt werden.³

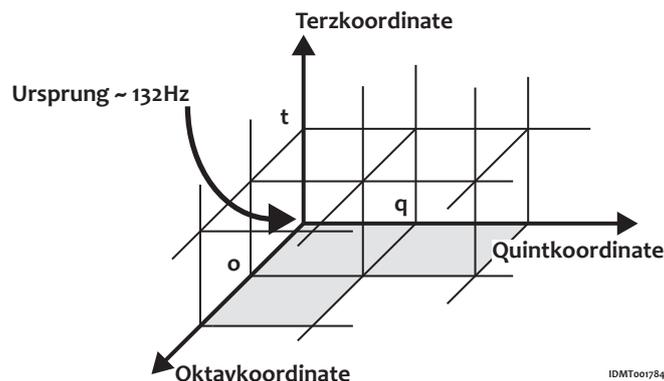


Abbildung 3.3 – Der Eulertonraum [109, S. 28]

3.4 Das Euler/Riemann/Tonnetz

Projiziert man alle Punkte des dreidimensionalen Eulerraumes in die zweidimensionale Terz-Quinte-Ebene, so erhält man das sogenannte Euler-Netz. Die Projektion kann als eine Abbildung aller Töne in den Bereich einer Oktave interpretiert werden. Durch diesen Schritt wird die in einem Ton enthaltene Oktavinformation vernachlässigt. Man redet in diesem Fall nicht mehr von Tönen, sondern von Tonigkeiten (siehe Abschnitt 2.3.3).

³ Ausgehend von einem periodischen Obertonspektrum stellen die Punkte 2^p die Oktav-Obertöne dar, die Punkte 3^s die Quint-Obertöne und die Punkte 5^r die Terzobertöne.

Die erste Achse des Euler-Netzes ordnet Tonigkeiten im Quintabstand an, die zweite im Terzabstand. Obwohl die ursprüngliche Intention des Netzes darin bestand, die reine Stimmung zu repräsentieren, entdeckten Musiktheoretiker wie Arthur von Oettingen oder Hugo Riemann (z.B. [133, S. 32ff]) das Netz zur Darstellung des Aufbaus und der Verwandtschaft von Dur- und Molldreiklängen. Wie in Abbildung 3.4 gesehen werden kann, können die Tonigkeiten im Euler-Netz zu dreieckigen Zellen verbunden werden. Die an den Eckpunkten einer solchen Zelle befindlichen Tonigkeiten bilden dabei einen Dur- oder einen Molldreiklang. Die drei in Abbildung 3.4 mit *P*, *V* und *G* bezeichneten Doppelpfeile stellen die Transformation des Akkordes C-Dur in dessen Parallel-, Variant- und Gegendreiklang dar (Abschnitt 5.3.4) und können durch Spiegelung des Akkorddreieckes an dessen eigenen Kanten erzeugt werden (vgl. [84], [101, S. 83]). Um die quintverwandten Dreiklänge eines gewählten Dreiklanges zu erhalten, muss das entsprechende Akkorddreieck in horizontaler Richtung verschoben werden. Eine weitere Eigenschaft des Tonnetzes ist, dass die Verwandtschaftsbeziehungen für Akkorde und Tonarten in gleicher Weise repräsentiert werden: Das diatonische Tonset (Abschnitt 2.6.1) belegt im Tonnetz einen zusammenhängenden Bereich, welcher in Abbildung 3.4 gestrichelt umrandet ist. Durch Verschieben dieses Bereiches nach schräg rechts unten (im Fall von Dur) oder oben (im Fall von Moll) selektiert man die Tonigkeiten der Varianttonart. Durch horizontales Verschieben des Tonartbereiches erreicht man die nächstgelegenen quintverwandten diatonischen Tonsets. An dieser Stelle wird bereits die Ähnlichkeit des Tonnetzes zum Vial/Weber/Schönbergnetz (Abschnitt 3.2) sichtbar: Fasst man in diesem Tonraum (Abbildung 3.2c) die Repräsentation von Dur- und Molltonart zusammen und stellt beide Modi an einem gemeinsamen Raumpunkt dar, so erhält man ein gedrehtes Tonnetz (vgl. [101, S. 99]). Somit wird deutlich, dass – wie im Tonnetz dargestellt – ein enger Zusammenhang zwischen Einzeltonigkeiten und Akkorden besteht, und Einzeltonigkeiten und Tonarten, wie im Vial/Weber/Schönbergnetz veranschaulicht.

Das Riemannnetz erlangte in den letzten Jahren wieder größere Bedeutung. Hier sollen nun einige wichtige Weiterentwicklungen und Anwendungen des Tonnetzes zusammengefasst werden:

- Gerald Balzano [9][10] zeigte, dass eine Kombination aus Groß- und Kleinterzkreisen eine räumliche Struktur ergibt, welche äquivalent zu Riemanns Tonnetz ist [101] (Abschnitt 3.5). Diese Erkenntnis machen sich Christopher Harte, Mark Sandler und Martin Gasser [58] in Form eines Algorithmus zunutze, der harmonische Änderungen in Audiosignalen detektiert (Abschnitt 3.5).
- Ein für die temperierte Stimmung adaptiertes Tonnetz ist die Basis von David Levins

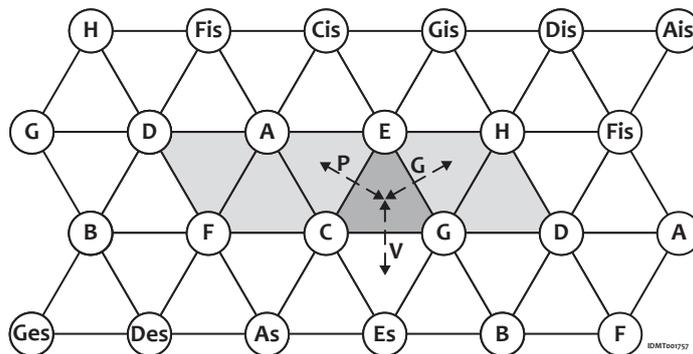


Abbildung 3.4 – Das Tonnetz, übernommen aus [49] und angepasst an aktuelle deutsche Begriffsgebung der Funktionstheorie [84]. Die Buchstaben *P*, *G* und *V* repräsentieren die Verwandtschaftsgrade Paralleldreiklang, Gegendreiklang und Variantdreiklang. Die graue Fläche markiert die leitereigenen Dreiecke des diatonischen Tonsets C-Dur (ohne verminderten Dreiklang)

[102] generalisierter Intervall- und Transformationstheorie (Neo-Riemannian Theory)⁴. Während Hugo Riemann [133] eher die statische „Funktion“ von Akkorden bezogen auf ein tonales Zentrum mit Hilfe des Tonnetzes im Blick hatte, fokussiert die Transformationstheorie auf die funktionelle Entwicklung eines Musikstückes durch Transformation eines Klanges in einen anderen [101, S. 215].

- Aline Honingh [69] zeigt, dass nicht nur das diatonische Tonset, sondern auch das tonale Material vieler anderer Skalen im Tonnetz zusammenhängende Bereiche belegen. Aus dieser Entdeckung leitet sie die Eigenschaft der Star-Konvexität als ein Maß für die Wohlgeformtheit tonaler Strukturen ab.
- Elaine Chew [20] leitet aus dem Tonnetz das sogenannte Spiral Array ab und entwickelt einen Algorithmus (Center-of-Effect-Generator CEG), der den harmonischen Schwerpunkt von Musikstücken im Spiral Array berechnet. Als Anwendung des Modells schlägt Chew einen Tonarterkennungsalgorithmus vor (Abschnitt 3.6).
- Richard Cohn [25] geht davon aus, dass die Bedeutung des Dur- und Molldreiklanges stark überbewertet wird und verallgemeinert deshalb das Euler/Riemann/Tonnetz. Die Hauptidee von Cohns Ansatz besteht darin, die Anordnung der Tonigkeiten nicht auf Quinten und großen Terzen festzulegen, sondern offen zu lassen.

⁴ Siehe auch Richard Cohn [25], [26], Edward Gollin [50] and Brian Hyer [71]

Bewertung:

Die Leistungsfähigkeit des Tonnetzes besteht in der logischen Systematisierung vieler musiktheoretischer Zusammenhänge rund um Dur- und Molldreiklänge. Eine besondere Eigenschaft des Tonnetzes ist dabei vor allem die Herstellung eines physikalischen Bezugs: Quinte und Terz als Basisintervalle des Netzes stellen eine Verbindung zwischen Frequenzverhältnissen von Tönen, Dur- und Molldreiklängen sowie Dur- und Molltonarten her (Ähnlichkeit zum Vial/Weber/Schönbergnetz). Eine Schwäche des Tonnetzes ist die mangelnde Trennung zwischen tonartbezogenen und tonartübergreifenden Aspekten. Das diatonische Tonset, welches primär für die abendländische Tonalität ist, geht nur als sekundäre Eigenschaft aus dem Netz hervor [101]. Es werden Modelle benötigt, welche die Diatonik und die Chromatik stärker voneinander abgrenzen. Zur Beschreibung musikpsychologischer Akkordähnlichkeiten (Abschnitt 5.3.4) ist das Tonnetz nicht geeignet. Die Wahrnehmung von Akkorden ist zum einen stark vom Kontext abhängig und kann nicht ohne einen solchen beschrieben werden. Zum anderen gibt es weitere Gesetze der Akkordwahrnehmung, die im Tonnetz nicht wiedergegeben werden. Dazu gehört z.B. der Quintfall: Die Wahrscheinlichkeit, dass ein Akkord von seinem im Quintenzirkel absteigend benachbarten Akkord gefolgt wird, ist deutlich höher als der umgekehrte Fall. Auch kadenzielle Zusammenhänge wie z.B. die hohe Wahrscheinlichkeit eines Folgens der Dominante auf die Subdominante werden im Tonnetz geometrisch nicht veranschaulicht.

3.5 Der Terztorus

Gerald Balzano [9][10] zeigte, dass eine Kombination aus Groß- und Kleinterzkreisen eine räumliche Struktur ergibt, welche äquivalent zu Riemanns Tonnetz ist [101] und als Terztorus bezeichnet wird (Abbildung 3.5b) [109, S. 104ff]. Die Bildung des Terztorus setzt die Annahme von Oktaväquivalenz (Abbildung 3.5a, gestrichelter Pfeil) und enharmonischer Äquivalenz (durchgezogene Pfeile) voraus.

Bewertung:

Der Terztorus veranschaulicht die Zirkularität des Tonnetzes in beiden Dimensionen, wenn man Oktav- und enharmonische Äquivalenz annimmt. Die toroidale Darstellung des Tonnetzes ist deshalb auch ein wichtiger Schritt für die Entwicklung tonnetz- und chromavektorbasierter Harmonieanalysealgorithmen wie z.B. dem von Harte, Sandler und Gasser [58]. Kritisch zu betrachten ist jedoch die Bedeutung von Balzanos Zerlegung

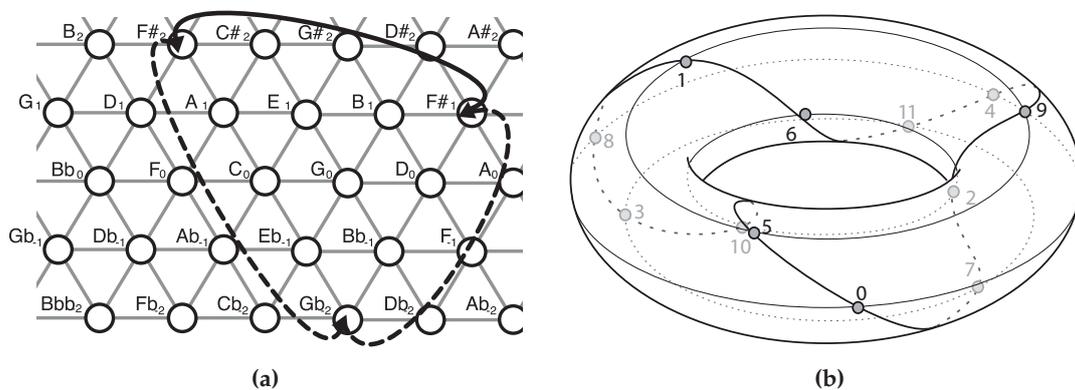


Abbildung 3.5 – Unter Annahme von Oktav- und enharmonischer Äquivalenz kann das Tonnetz als ein Torus dargestellt werden (angepasst aus [58]).

des Tonnetzes in eine zyklische Groß- und Kleinterzkomponente. Es ist fraglich, ob diese beiden Komponenten wahrnehmungspsychologisch relevante Achsen darstellen und damit jeweils für die musikalische Ähnlichkeitsanalyse interessante Features repräsentieren. Entsprechend Lerdahl [101, S. 44ff] haben Groß- und Kleinterzzyklen erst relativ spät (im 19. Jahrhundert) mit der Entstehung chromatischer Harmonievorgänge Bedeutung gewonnen und stellen eher eine Weiterentwicklung der aus dem Quintzyklus heraus entstandenen diatonischen Harmonik dar als deren strukturelle Grundlage.

3.6 Elaine Chews Spiral Array

Elaine Chews Spiral Array [20] entspricht vom Aufbau her dem Terztorus (Abschnitt 3.5). Abbildung 3.6 stellt eine dreidimensionale Verzerrung des eigentlich vierdimensionalen Torus dar, bei dem es sich genau genommen um eine Spirale handelt, die sich in der vierten Dimension schließt. Verzichtet man auf die Darstellung der Geschlossenheit des Terztorus in der vierten Dimension und stellt diesen als sich nicht schließende Spirale dar, so erhält man Elaine Chews Spiral Array (Abbildung 3.6). Dies bedeutet, dass alle bisher diskutierten Eigenschaften des Tonnetzes auch für das Spiral Array gelten. Das in Chews Dissertation adressierte Problem des Tonnetzes besteht darin, dass Dreiklänge und Tonarten im Tonnetz nicht durch einen einzelnen eindeutig bestimmten Punkt, sondern als durch mehrere Punkte definierte Flächen bestimmt sind. Chew bestimmt deshalb mit Hilfe von Methoden aus der Unternehmensforschung (Operations Research) Punkte innerhalb der von einem Dreiklang im Tonnetz aufgespannten Dreiecke (Abbildung 3.6, hellgraue Fläche), wobei jeder der so gefundenen geometrischen Punkte einen Dur- oder Molldreiklang repräsentiert. Im nächsten Schritt verbindet Chew die zuvor gefundenen

Akkordschwerpunkte zu neuen Dreiecken (Abbildung 3.6, dunkelgraue Fläche), welche nun Tonarten repräsentieren. Das eine Tonart bildende Dreieck wird dabei durch die Schwerpunkte der Hauptdreiklänge (Abschnitt 5.3.1) der Tonart gebildet, das sind die Tonika, die Subdominante und die Dominante. Analog zu den Dreiklangsschwerpunkten berechnet Chew schließlich Tonartschwerpunkte (Abbildung 3.6, sternförmige Punkte innerhalb des dunkleren Dreiecks). Als Evaluation ihres Modells schlägt Elaine Chew einen Tonarterkennungsalgorithmus vor, welcher den geometrischen Schwerpunkt eines Tonsets im Spiral Array berechnet und auf Basis der nächst gelegenen Tonartpunkte eine Entscheidung trifft. Weitere Details des Tonarterkennungsalgorithmus sind in Abschnitt 9.1.2 zu finden.

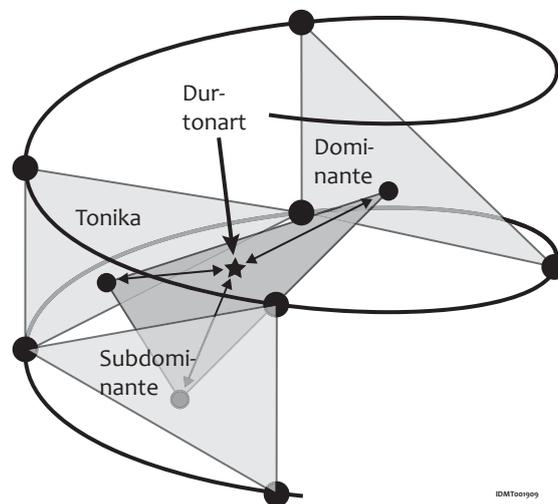


Abbildung 3.6 – Elaine Chews Spiral Array für Durtonarten (nach [20])

Bewertung:

Chews Beitrag besteht zum einen in der spiralförmigen Darstellung des Terztorus. Durch diese Darstellung werden viele Zusammenhänge einfacher ersichtlich. Weiterhin ist die Idee bahnbrechend, die Töne eines Musikstückes in ein geometrisches Raummodell abzubilden, darauf aufbauend Schwerpunkte zu berechnen und auf dieser Basis Tonarterkennung zu betreiben. Die Schwerpunkte von Akkorden und Tonarten mit Hilfe von Methoden der Unternehmensforschung zu bestimmen ist interessant. Jedoch wäre hier ein Vergleich der optimierten geometrischen Tonart- und Akkordschwerpunkte sowohl mit einfachen als auch mit Trainingsdaten gewonnenen Schwerpunkten wünschenswert.

3.7 Fred Lerdahls Tonal Pitch Space

Das wohl umfassendste und von einem theoretischen Standpunkt aus gesehen leistungsfähigste aller momentan bestehenden Tonraummodelle ist Fred Lerdahls „Tonal Pitch Space“ [101]. Dabei handelt es sich nicht um einen einzelnen Tonraum, sondern ein Framework aus mehreren zusammenhängenden Tonräumen. Da dieses Framework viele wichtige Ansätze für die harmonieorientierte semantische Metadatenanalyse enthält, sollen zunächst die wichtigsten Eigenschaften kurz zusammengefasst und anschließend detailliert beschrieben werden.

Überblick

- Fred Lerdahls Modell besteht aus drei unterschiedlichen Tonraummodellen: Das erste Modell ist der sogenannte *Basic Space*, welcher die Eigenschaften *einzelner Tonigkeiten* in Bezug auf eine bestimmte Tonleiter und einen bestimmten Akkord auf fünf verschiedenen tonalen Ebenen beschreibt. Zu diesen Ebenen gehören die „Oktavebene“, die „Quintebene“, die „Akkordebene“, die „Tonleiterebene“ und die „chromatische Ebene“. Der Basic Space ist an empirischen Ergebnissen Krumhansls zur Wahrnehmung von Tonigkeiten angelehnt ([81], Abschnitt 2.7.1) und spiegelt damit wichtige wahrnehmungspsychologische Zusammenhänge wider.
- Der zweite Tonraum ist der sogenannte *Chordal Space*, welcher *Akkorde* einer gegebenen Tonart in einem zweidimensionalen Modell organisiert. Den musiktheoretischen Grundlagen entsprechend werden Akkorde entlang der ersten Achse in Quinten und entlang der zweiten Achse auf Basis gemeinsamer Tonigkeiten angeordnet.⁵ Der Chordal Space enthält nur leitereigene Tonigkeiten und zieht damit eine klare Grenze zwischen leitereigenem und leiterfremdem Tonmaterial.
- Der dritte Tonraum ist der sogenannte *Regional Space*, welcher *Tonarten* repräsentiert und im Wesentlichen dem Vial/Weber/Schönbergnetz entspricht, welches in Abschnitt 3.2 bereits ausführlich erläutert wurde.
- Eine Besonderheit von Lerdahls Modell ist die *baumförmige Verknüpfung* der drei zuvor beschriebenen Tonraummodelle. Wie in Abbildung 3.7 zu sehen, stellt der Regional Space die Wurzel dieses Baumes dar. Der Regional Space organisiert 24 Dur- und Molltonarten. Jede dieser 24 Dur- und Molltonarten wiederum wird

⁵ D.h. Akkorde, die zwei gemeinsame Tonigkeiten besitzen, werden benachbart angeordnet.

durch einen dafür konfigurierten Chordal Space repräsentiert.⁶ Jeder dieser Chordal Spaces enthält 7 räumlich angeordnete leitereigene Dur- und Molldreiklänge, welche jeweils durch einen dafür konfigurierten Basic Space repräsentiert werden. Letzterer besteht aus den bereits benannten fünf hierarchischen Ebenen.

- Die baumförmige Verknüpfung der Teilräume von Lerdahls Tonraum ermöglicht die Berechnung der *perzeptuellen Nähe* von Tonigkeiten und Akkorden über Tonartgrenzen hinweg. Die Nähe zweier tonaler Elemente ergibt sich dabei aus dem Abstand der Elemente im Tonraum.
- Um perzeptuelle Asymmetrien zu modellieren (Abschnitt 2.7.4), führt Lerdahl zusätzlich noch ein *Kräftemodell* ein. Dieses ermöglicht die Berechnung der *harmonischen Spannung* zwischen Akkorden sowie die Berechnung von *tonalen Anziehungskräften* zwischen Tonigkeiten und Akkorden. Mit Hilfe des Kräftemodells ist Lerdahl in der Lage, z.B. das Auflösungsbestreben des Leittones in den Grundton zu modellieren.

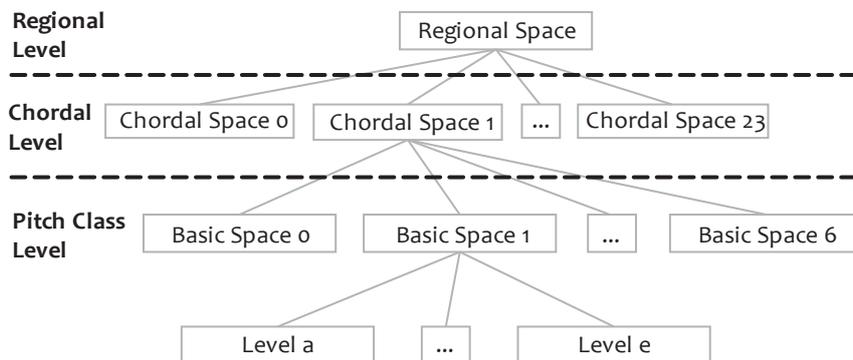


Abbildung 3.7 – Die baumförmige Verknüpfung der Teilräume von Fred Lerdahls Tonal Pitch Space

Nachdem ein Überblick über Lerdahls Tonal Pitch Space gegeben wurde, erfolgt nun eine detailliertere Beschreibung der einzelnen Aspekte des Modells.

Der Basic Space

Ausgangspunkt von Lerdahls „Tonal Pitch Space“ ist der sogenannte *Basic Space*. Die Anordnung der Töne in diesem Raum ist an eine Theorie zur Tonwahrnehmung von

⁶ D.h. es gibt 24 Chordal Spaces für 24 Tonarten.

Diana Deutschs und John Feroes [33] und Ergebnisse Carol L. Krumhansls zur Ähnlichkeitswahrnehmung von Tonigkeiten in einer gegebenen Tonart (Abschnitt 2.7) angelehnt. Nach dieser Theorie strukturieren Hörer tonale Sequenzen wie hierarchisch organisierte Alphabete. Das übergeordnete Alphabet hierbei ist die Oktave, dann kommen Dreiklänge und Septakkorde, dann die diatonische Tonleiter und danach die chromatische Tonleiter. Jede Ebene „schmückt den Inhalt der darüberliegenden“ Ebene aus. Abbildung 3.8 zeigt den Basic Space. Entsprechend den vorherigen Ausführungen besteht dieser aus fünf Ebenen: a) Oktavebene, b) Quintebene, c) Akkordebene, d) diatonische Ebene und e) chromatische Ebene. Dabei gilt: Je höher eine Ebene ist, in der sich eine Tonigkeit befindet, desto höher ist die Stabilität der Tonigkeit. Stabile Tonigkeiten wiederholen sich in darunter liegenden Ebenen. Grundlage der im Basic Space gezeigten Hierarchie sind die Oktave und die Quinte als konsonanteste Intervalle und nicht die Chromatik (letzte Ebene). Der Basic Space ermöglicht es, die *Instabilität von Tonigkeiten* in Bezug auf *einen gegebenen Akkord* zu berechnen. Sie ergibt sich laut Lerdahl aus der Anzahl der horizontalen und vertikalen Schritte, die minimal notwendig sind, um die höchste Tonigkeit in der Hierarchieebene zu erreichen.⁷ Der Basic Space modelliert das Verhalten von Tonigkeiten bezogen auf einen bestimmten Akkord. Um weitere Akkorde abbilden zu können, muss der Basic Space auf den Ebenen *a* bis *c* umkonfiguriert werden. Die Umkonfiguration erfolgt nicht auf Basis einer fest definierten Anzahl von Halbtonschritten (z.B. sieben Halbtonschritten für eine Quinte), sondern auf Basis einer Anzahl von Tonleiterschritten, welche in der Ebene *d* repräsentiert sind. Um den Basic Space in einen quintverwandten Akkord umzukonfigurieren, müssen alle Tonigkeiten auf den Ebenen *a* bis *c* um Schritte bezogen auf Ebene *d* verschoben werden. Die Entscheidung, die Umkonfiguration auf Basis von Ebene *d* durchzuführen, ist sehr sinnvoll, da dadurch der diatonische Bereich nicht verlassen wird, d.h. es werden nur leitereigene Akkorde erzeugt.

level a:	C												(C)
level b:	C						G						(C)
level c:	C			E			G						(C)
level d:	C		D		E	F		G		A		H	(C)
level e:	C	Des	D	Es	E	F	Fis	G	As	A	B	H	(C)

IDMT000782

Abbildung 3.8 – Der Basic Space für den Akkord C-Dur [101, S. 47]

⁷ Im Beispiel von Abbildung 3.8 besitzt die Tonigkeit C die geringste Instabilität (0). Die Tonigkeit D dagegen besitzt die Instabilität 4, da ein horizontaler und vier vertikale Schritte erforderlich sind, um die höchste Tonigkeit *c* zu erreichen.

Der Chordal Space

Zur Bestimmung der Zusammengehörigkeit von Akkorden schlägt Lerdahl vor, zwei Kriterien heranzuziehen: 1.) die Quintverwandtschaft von Akkorden und 2.) die Anzahl der gemeinsamen Töne. Ordnet man Akkorde einer Tonart in dieser Art an, dann erhält man die zwei in Abbildung 3.9 gezeigten Kreissysteme: Das erste Kreissystem ist der „*Diatonic Circle of Fifths*“ (Abbildung 3.9a) und das zweite Kreissystem ist der „*Common Tone Circle*“ (Abbildung 3.9b). Analog zum Terztorus (Abschnitt 3.5) ergeben diese beiden Kreissysteme einen vierdimensionalen Torus. Die Abbildung der beiden Winkeldimensionen in ein XY-Koordinatensystem ist in Abbildung 3.10 zu sehen. Im Chordal Space schreiten die Akkorde in der ersten Winkeldimension in Quinten fort und in der zweiten Winkeldimension in großen und kleinen Terzen. Anwendung des Chordal Spaces ist die Berechnung des perzeptuellen Abstandes von Akkorden innerhalb einer Tonart (siehe unten).

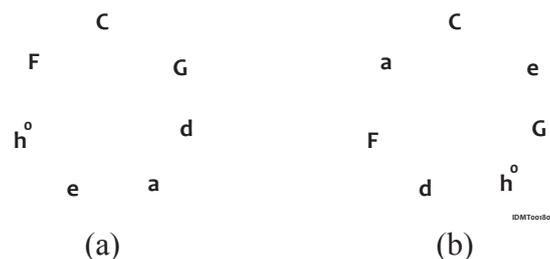


Abbildung 3.9 – (a) Der Diatonic Circle of Fifths für C-Dur (nach [101, S. 55]) ordnet Tonigkeiten in Quintverwandtschaft an. (b) Der Common Tone Circle für C-Dur (nach [101, S. 56]) ordnet die Grundtöne leitereigener Dreiklänge (im Beispiel C-Dur) entsprechend der Anzahl gemeinsamer Tonigkeiten auf einem Kreis an.

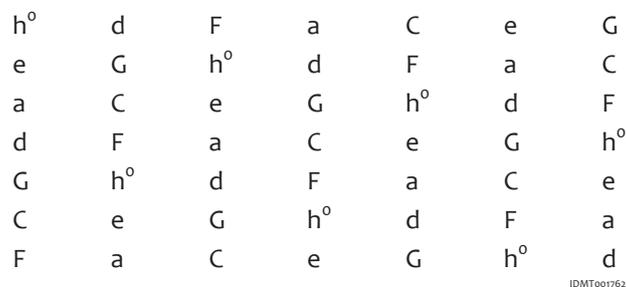


Abbildung 3.10 – Der Chordal Space, eine zweidimensionale Darstellung der Kombinationen aus Abbildung 3.9a und Abbildung 3.9b (nach [101])

Der Regional Space

Der im vorherigen Abschnitt beschriebene Chordal Space ermöglicht nur das Vergleichen von Akkorden einer konkreten Tonart. Um auch den wahrgenommenen Abstand von Akkorden über Tonartgrenzen hinaus zu modellieren, führt Lerdahl den Regional Space ein. Lerdahl greift dabei auf das Vial/Weber/Schönbergnetz (Abbildung 3.2) zurück, da dieses in weiten Teilen sowohl theoretisch (Vial/Weber/Schönberg) als auch empirisch (Krumhansl) untermauert ist. Ähnlich wie jedes Element des Chordal Spaces mit einem zugehörigen Basic Space verknüpft ist, ist auch jedes Element des Regional Space mit einem zugehörigen Chordal Space verknüpft. Dadurch ist die Berechnung des perzeptuellen Abstandes von Akkorden aus unterschiedlichen Tonarten möglich.⁸ Der Abstand ergibt sich dabei aus der Kombination der Entfernung der beiden Tonarten im Regional Space, zu denen die Akkorde gehören, und dem Abstand der Akkorde zur Tonika der zugehörigen Chordal Spaces (Chord/Region Distance Rule [101, S. 70]).

Perzeptuelle Nähe von Tonigkeiten, Akkorden und Tonarten

Im Abschnitt 2.7 wurde anhand von Experimenten von Carol Krumhansl gezeigt, dass die Einzeltöne, Akkorde und Tonarten in ihrer Ähnlichkeit unterschiedlich wahrgenommen werden. Da die Nähe zweier Tonigkeiten laut Krumhansl [81] stark von der Existenz eines tonalen Zentrums abhängt, d.h. einer Tonika, berechnet Lerdahl die Abstände einzelner Tonigkeiten im Basic Space. Dieser setzt jede Tonigkeit in Bezug zu einer gegebenen Tonika. Der Abstand von Akkorden ergibt sich wie bereits zuvor erwähnt aus einer Kombination der Abstände der zugehörigen Tonarten und der Position der Akkorde im Chordal Space. Für die Abstandsberechnung wendet Lerdahl das „Prinzip des kürzesten Pfades“ an, nach welchem die kürzest mögliche Verbindung zweier Elemente in den jeweiligen Tonräumen der wahrgenommenen Ähnlichkeit entspricht.⁹ Die Verbindung zweier Elemente im Tonraum wird dabei durch horizontale, vertikale oder teilweise auch diagonale¹⁰ Pfade hergestellt. Die besondere Leistungsfähigkeit von Lerdahls Modell

⁸ Z.B. kann der Abstand der Akkorde *F*-Dur und *G*-Dur einmal für den Fall berechnet werden, dass diese beide Bestandteil der Tonart *C*-Dur sind oder in einem anderen Fall Bestandteil der Tonart *B*-Dur und der Tonart *G*-Dur. Die wahrgenommene Nähe der beiden Akkorde ist in beiden Fällen sehr verschieden (vgl. Carol L. Krumhansl [82, Abbildung 8]).

⁹ Lerdahl bezieht sich dabei auf das Gestaltprinzip der Prägnanz, nachdem die einfachste mögliche Interpretation eines Musikstückes der tatsächlichen Wahrnehmung entspricht. Dieser Gedanke entspricht auch Hugo Riemanns Idee von der „Ökonomie des Vorstellens“ [133, S. 24].

¹⁰ Diagonale Bewegungen im Regional Space werden erlaubt, wenn die Tonika der Zieltonart auch leitenderer Klang der Ausgangstonart ist.

besteht deshalb darin, mit Hilfe der zuvor beschriebenen baumförmigen Verknüpfung tonart- bzw. akkordbezogener Tonraummodelle den Abstand von Elementen aus beliebigen Tonarten berechnen zu können, was schon im Abschnitt zum „Chordal Space“ angesprochen wurde.

Einordnung von Spannungskräften in den Tonraum

Neben der wahrgenommenen Nähe zweier Elemente schlägt Lerdahl ein Modell zur Berechnung der wahrgenommenen harmonischen Spannung vor.¹¹ Bei der *harmonischen Spannung* unterscheidet Lerdahl zwischen drei unterschiedlichen Formen, nämlich der sequentiellen Spannung (Sequential Tension), der hierarchischen Spannung (Hierarchical Tension) und der Oberflächenspannung (Surface Tension). Die *sequentielle Spannung* zwischen zwei Akkorden ergibt sich dabei aus ihrer Entfernung im Tonraum. Für die Berechnung der *hierarchischen Spannung* wird das Musikstück zuvor den vordefinierter Regeln entsprechend segmentiert und in eine baumförmige Struktur (Prolongational Tree) „geparst“. Die hierarchische Spannung eines Akkordes ergibt sich anschließend durch Addieren der räumlichen Abstände in der Hierarchie vertikal benachbarter Elemente [101, S. 143] und berechnet damit die Spannung eines Akkordes in Bezug auf eine Art Referenzakkord. Mit der dritten Form der harmonischen Spannung, der *Oberflächenspannung*, bezieht Lerdahl weiterhin den Einfluss der Akkordumkehrung sowie akkordfremder Tonigkeiten ein (Vorhalte, Durchgangsnoten, Charakterdissonanzen). Letztere erhöhen ebenfalls die zu einem Zeitpunkt hörbare Spannung.

Einordnung von Anziehungskräften in den Tonraum

In der musikalischen Praxis unterscheiden sich sowohl die Übergangswahrscheinlichkeiten einer Tonkombination A in eine Tonkombination B oft deutlich von der umgekehrten Reihenfolge. So ist z.B. der Quintfall von Akkorden deutlich häufiger anzutreffen als der Quintstieg. Durch räumliche Abstände kann dieses Phänomen nicht modelliert werden (vgl. [101, S. 119ff]). Lerdahl löst dieses Problem, indem er neben den zuvor besprochenen Spannungskräften auch noch harmonische *Anziehungskräfte* (Harmonic Attraction) in den Tonraum einordnet. Harmonische Spannung hängt mit der Tendenz von instabilen Tonigkeiten oder Akkorden zusammen, unmittelbar von stabileren Tonigkeiten oder Akkorden gefolgt zu werden. Demnach wirken stabile Tonigkeiten als kognitiver Referenzpunkt oder auch Anker [101, S. 161], auf den instabile Tonigkeiten hinstreben. Bei der

¹¹ Neben der hier betrachteten harmoniebedingten Spannung gibt es noch weitere rhythmus-, dynamik-, melodie- oder auch klangfarbenbedingte Spannungen [101, S. 143]. Diese werden hier jedoch nicht betrachtet.

harmonischen Anziehung wiederum unterscheidet Lerdahl die melodische Anziehung (Melodic Attraction), die Stimmführungsanziehung (Voice-Leading Attraction) sowie die harmonische Anziehung (Harmonic Attraction). Für die Berechnung der *melodischen Spannung* [101, S. 161ff] schlägt Lerdahl vor, zum einen den Abstand zweier Tonigkeiten in Halbtonschritten einzubeziehen¹² und zum anderen jeder Tonigkeit eine Anziehungskraft (Anchoring Power) zuzuordnen, welche sich aus der Position der Tonigkeit im Basic Space ergibt. Je höher die Ebene ist, in der eine Tonigkeit im Basic Space vertreten ist, desto höher ist die Anziehungskraft der jeweiligen Tonigkeit. Die *Stimmführungsanziehung* [101, S. 173] berechnet die Anziehung zwischen zwei Akkorden. Sie ergibt sich durch Addition der Anziehungskräfte der Einzelstimmen dieser Akkorde. Die *Harmonische Anziehung* zwischen zwei Akkorden ergibt sich aus dem Quotienten aus Stimmführungsanziehung und der Entfernung der beiden Akkorde im Tonraum. D.h. je größer der Abstand zweier Akkorde im Tonraum, desto kleiner ist die harmonische Anziehung.

Weiterführende Aspekte

- Da viele musikalische Vorgänge stark von der zu einem Zeitpunkt etablierten Tonart abhängen, ist die Frage der Tonarterkennung (Tonic Finding) sehr relevant. Im Kapitel „Prolongational Functions“ stellt Lerdahl einen Ansatz zur Tonarterkennung vor, welcher auch harmonisch entfernte Akkorde mit einbezieht und dadurch für erweiterte Tonarterkennungsverfahren interessant ist.
- Wichtig beim perzeptuellen Vergleich von Musikstücken ist die Abschätzung der kontextabhängigen Momentanwirkung eines Akkordes. Letztere wird musiktheoretisch durch die Funktion des jeweiligen Akkordes beschrieben. Lerdahl diskutiert dazu im Kapitel „Prolongational Functions“ ein gegenüber Hugo Riemann [134] erweitertes/verändertes Funktionsmodell, das in die Entwicklung funktionsbezogener Audioanalysealgorithmen einbezogen werden sollte.
- Der Schwerpunkt von Lerdahls Tonal Pitch Space liegt auf der Modellierung diatonischer Harmonik. Im 19. Jahrhundert haben sich jedoch neue Modulations- und Ausweichungsformen herausgebildet, welche sehr kurze Verbindungen zwischen weit entfernten Tonarten herstellen konnten. In einem Kapitel „Chromatic Tonal Spaces“ erweitert Lerdahl sein Modell um derartige Aspekte. Die im 20. Jahrhundert entstandene atonale Harmonik ist Bestandteil eines weiteren Kapitels „Atonal Structures“.

¹² Je kleiner der Halbtonabstand zwischen zwei unterschiedlichen Tonigkeiten, desto größer deren Anziehung

Bewertung:

Mit dem Tonal Pitch Space stellt Lerdahl ein umfassendes Tonraummodell zur Verfügung, welches verschiedene mit anderen Tonraummodellen bisher noch nicht gelöste Probleme adressiert. Dazu gehören die Berücksichtigung der Unterschiede in der Wahrnehmung der Ähnlichkeit von Einzeltönen, Akkorden und Tonarten in Form dreier unterschiedlicher Tonraummodelle. Durch den Basic Space stellt Lerdahl darüber hinaus einen Zusammenhang zwischen den unterschiedlichen musikalischen Ebenen „Harmonik“, „Diatonik“ und „Chromatik“ her. Die Modellierung perzeptueller Asymmetrien, laut Krumhansl [82] eine Schwäche existierender Tonraummodelle, löst Lerdahl durch die Einordnung eines Kräftemodells, das sowohl Spannungskräfte als auch tonale Anziehungskräfte berechnen kann. Eine weitere Schwäche bisheriger Tonräume, nämlich tonartübergreifende und tonartbezogene harmonische Vorgänge strukturell nur unzureichend abzugrenzen, löst Lerdahl durch die Bereitstellung speziell für einen gegebenen Kontext konfigurierter Spezialmodelle (Basic und Chordal Space).¹³ Insgesamt muss gesagt werden, dass Lerdahls Modell viele für den harmoniebasierten perzeptuellen Vergleich von Musikstücken interessante Parameter beinhaltet. Da das Modell jedoch für die Analyse symbolischen Audios ausgelegt ist, müssen Verfahren entwickelt werden, welche die vorgeschlagenen Ansätze für die Analyse von echtem Audio einsatzfähig machen. Problematisch hierbei ist, dass Lerdahl in seinen Analysen stets von eindeutig bestimmten Akkorden und Tonarten ausgeht, was jedoch in der Praxis nicht der Fall ist. Die Existenz leistungsfähiger Akkord- und Tonarterkennungsalgorithmen ist demnach eine Grundvoraussetzung, um Musikstücke anhand der von Lerdahl vorgeschlagenen Parameter zu vergleichen. Um Lerdahls Kräfte- und Funktionsmodell in vollem Umfang zu nutzen, sind zusätzlich noch Algorithmen notwendig, welche in der Lage sind, Audiosignale in die darin verborgene baumförmige Hierarchie (Time Span Tree, Prolongational Tree) zu parsen. Ansätze für symbolisches Audio hierfür sind z.B. in [55] zu finden, aufgrund der Komplexität von realem Audio (vgl. Kapitel 7) müssen an dieser Stelle jedoch starke Vereinfachungen vorgenommen werden. Eine multimediale Umsetzung von Lerdahls Arbeit ist in [167] zu finden.

¹³ Die Verwendung von nicht leitereigenem Material wie z.B. Akkorden aus anderen Tonarten stellt in der Regel einen deutlich wahrnehmbaren Ausbruch aus der Standard-Tonalität dar. In den bisher diskutierten Tonraummodellen existieren keine tonalen Grenzen, die „überwunden“ werden müssen, um auf leiterfremdes Material zugreifen zu können. Lerdahl löst dieses Problem auf zwei Ebenen: Zum einen stellt er mit dem Basic Space einen Tonraum bereit, der klar auf einen Akkord bezogen ist, und zum zweiten mit dem Chordal Space einen weiteren. Der wiederum ist auf eine Tonart bezogen. Das Ausweichen auf leiterfremdes tonales Material kann in Lerdahls Modell nur durch eine Umkonfiguration des Tonraumes modelliert werden.

3.8 Das harmonische Heptagon/ Der Kadenzkreis

In seinem Buch „What is Music“ [34] stellt Philip Dorell einen als „Harmonisches Heptagon“ (Harmonic Heptagon) bezeichneten Tonraum vor (Abbildung 3.11a), welcher vom Aufbau her Fred Lerdahls Common Tone Circle (Abbildung 3.9a) entspricht. Das harmonische Heptagon ordnet die sieben Tonigkeiten eines diatonischen Tonsets in abwechselnd kleinen und großen Terzen an. Dorell verwendet das Harmonische Heptagon, um das syntonische Komma¹⁴ zu erklären: Stimmt man ausgehend von einer gegebenen Tonigkeit entsprechend dem Harmonischen Heptagon reine große und kleine Terzen, so unterscheiden sich die Frequenzen von Anfangs- und Endtonigkeit um das syntonische Komma. Weiterhin stellt Dorell heraus, dass die Tonigkeiten aller aus einem diatonischen Tonset heraus erzeugbaren Dur- und Molldreiklänge sowie der verminderte Dreiklang in Benachbarung liegen und dass der Raum eine innere Spiegelsymmetrie besitzt. Eine weitere von Dorell bemerkte Besonderheit ist die Symmetrie der Tonikadreiklänge C, E, G und A, C, E sowie die diesen Dreiklängen gegenüberliegende Tonigkeit D.

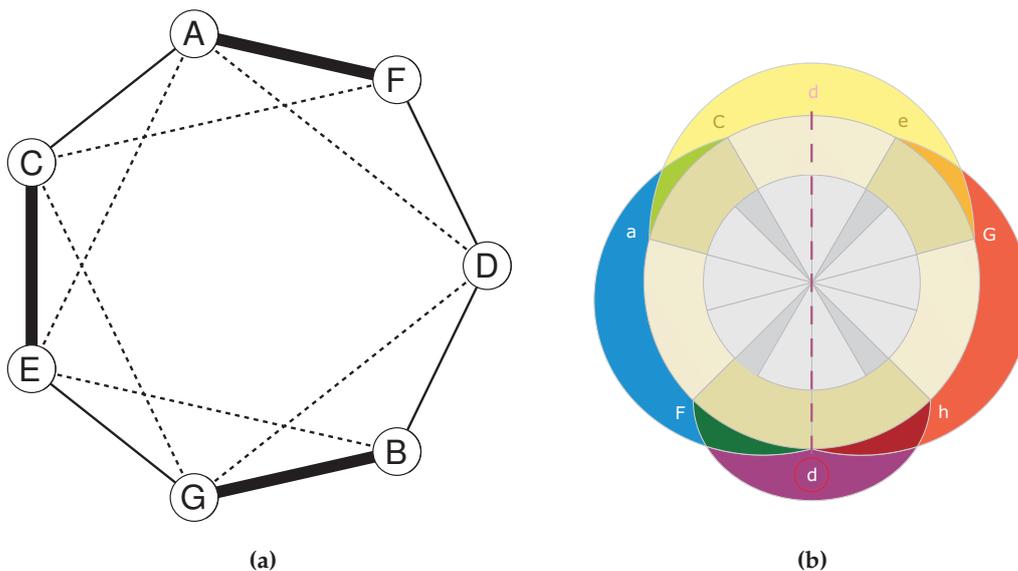


Abbildung 3.11 – (a) Das Harmonische Heptagon [34], (b) Der Kadenzkreis [47]

Nahezu zeitgleich zu Dorell veröffentlichte David Gatzsche [47] den in Abbildung 3.11b gezeigten Kadenzkreis, welcher identisch zum Harmonic Heptagon ist. Unterschiede von Gatzsches und Dorells Räumen sind die Winkelorientierung des Koordinatensystems sowie die Ausrichtung des Kreises, welche im Fall des Kadenzkreises an den Quintenzirkel

¹⁴ (siehe Abschnitt 2.6.3)

angelehnt sind. Gatzsches besonderer Beitrag war die Einführung der *Symmetrieachse* und des *Symmetrietones*. Im Fall der Tonart C-Dur ist dies der auch von Dorell herausgehobene Ton d. Gatzsche zeigte, dass sich zur Linken der Symmetrieachse die subdominantischen Dreiklänge des gewählten diatonischen Tonsets ($d-F-a$, $F-a-C$), zur Rechten die dominantischen Dreiklänge ($e-G-h$, $G-h-D$) und im oberen Bereich des Kreises die Tonikadreiklänge befinden ($a-C-e$, $C-e-G$). Gegenüber dem Tonikabereich liegt der verminderte Dreiklang ($h-d-F$), welcher von Gatzsche als Symmetrieakkord bezeichnet wird. Sowohl Symmetrieakkord ($h-d-F$) als auch Symmetrieton (d) verbinden die gegenüberliegenden funktionellen Pole Dominante und Subdominante. Für Gatzsche ist dies eine mögliche Erklärung für das hohe Auflösungsbestreben des verminderten Dreiklanges und damit auch des Dominantseptakkordes in Richtung Tonika. Das Ziel des Kadenzkreises bestand in der Bereitstellung eines pädagogischen Hilfsmittels zur Vermittlung des Aufbaus von Dur- und Molldreiklängen sowie funktionstheoretischer Zusammenhänge. Dazu entwickelte Gatzsche das in 3.11b gezeigte Farbsystem: In Anlehnung an Aussagen des französischen Musiktheoretikers und Komponisten Jean-Philippe Rameau, welcher tonale Bewegungen in Richtung Subdominante mit Kälte und Bewegungen in Richtung Dominante mit Wärme charakterisierte, wurde der Subdominantbereich mit den Farben blau, der Tonikabereich mit der Farbe gelb und der Dominantbereich mit der Farbe rot eingefärbt. Der spannungsgeladene verminderte Dreiklang wurde mit violett eingefärbt.

Bewertung:

Dorells Harmonic Heptagon bzw. Gatzsches Kadenzkreis bringen viele musiktheoretische Eigenschaften in Form eines sehr einfachen Modells zum Vorschein. Die Benachbarung der Tonigkeiten von leitereigenen Dreiklängen sowie die in den Kreisen zu Tage tretenden funktionellen Zusammenhänge machen den Tonraum interessant sowohl für die semantische Metadatenanalyse als auch für die tonraumbasierte Synthese von Musik. Jedoch müssen hier Ansätze entwickelt werden, Musiksignale in die entsprechenden geometrischen Modelle abzubilden. Für die Synthese von Musik ist die Frage, inwiefern derartige tonigkeitsorientierte Modelle um eine Tonhöhenkomponente erweitert werden können.

3.9 Generalized Voice-Leading Spaces

Clifton Callender, Ian Quinn und Dmitry Tymoczko veröffentlichten 2008 in der SCIENCE [19] einen Artikel über allgemeine Stimmführungsräume (Generalized Voice-Leading

Spaces). Inspiriert durch Ansätze aus der Musical Pitch-class Set Theory¹⁵ (musikalische Tonigkeitsmengenlehre) schlagen sie vor, Ton- und Akkordfolgen unter *Vernachlässigung* der Informationen Oktavierung, Permutation, Transposition und Änderung der Kardinalität zu klassifizieren. Darauf leiten sie fünf verschiedene Äquivalenzrelationen ab: Die *Oktaväquivalenz* definiert, dass zwei Töne im Abstand einer oder mehrerer Oktaven äquivalent sind. Die *Permutationsäquivalenz* definiert, dass verschiedene Umkehrungen eines Akkordes äquivalent sind. Die *Invertierungsäquivalenz* setzt Akkorde gleich, die durch Invertierung (Spiegelung am Ursprung des Koordinatensystems) ineinander übergeführt werden können. Die *Transpositionsäquivalenz* definiert, dass ein Akkord und dessen Transpositionen äquivalent sind. Und die *Kardinalitätsäquivalenz* besagt, dass ein Akkord, bei dem Töne verdoppelt werden, äquivalent zum Ausgangsakkord ist. Um diese Dinge geometrisch zu modellieren schlagen Callender, Quinn und Tymoczko vor, einen n -dimensionalen Raum R^n als Ausgangsbasis zu verwenden und dann je nach zu analysierender Äquivalenzrelation nach und nach in entsprechende Quotientenräume¹⁶ umzuformen. Die Anzahl der Dimensionen des Ausgangsraumes entspricht der Anzahl der Stimmen des Musikstückes. Zur Analyse eines vierstimmigen Satzes würde der Ausgangsraum also vier Dimensionen besitzen. Die Einteilung der Achsen des Ausgangsraumes erfolgt in Halbtonschritten. In Abbildung 3.12a ist der zweidimensionale Ausgangsraum für ein zweistimmiges Musikstück dargestellt. Insgesamt schlagen die Autoren vor, folgende Transformationen auf diesen Ausgangsraum anzuwenden:

- Bei Annahme von *Oktaväquivalenz* ist im R^n in allen Dimensionen jeder zwölfte Halbtonpunkt identisch. Für den in Abbildung 3.12 dargestellten Raum R^2 bedeutet dies, dass die gegenüberliegenden dick gestrichelt bzw. gepunktet hervorgehobenen Seiten identisch sind und deshalb verbunden werden können. Das Ergebnis ist ein n -Torus.
- *Permutationsäquivalenz* bedeutet, dass alle Punkte in R^n identisch sind, die durch Spiegelung an der den Raum ansteigend diagonal schneidenden Hyperebene¹⁷ in-

¹⁵ Die Pitch-class Set Theory wurde entwickelt, um ein generelles theoretisches Framework zur Beschreibung der Tonorganisation in posttonaler Musik bereitzustellen [87]. Zu herausragenden Vertretern dieses Gebietes gehören unter anderem Milton Babbitt, Allen Forte [45], John Rahn [130] oder auch David Lewin [102]. Modelliert man ein tonales Ereignis als eine Menge von Tönen und ein Musikstück als eine Folge solcher Mengen, so ist es möglich, die Beziehungen zwischen Tonsets mit Methoden der mathematischen Mengenlehre zu beschreiben.

¹⁶ Ein Quotientenraum wird durch „Aneinanderheften“ bestimmter Punkte eines Elternraumes gebildet. Die Art und Weise dieses Prozesses hängt von einer Menge von Operationen ab [19, Supporting Material].

¹⁷ Eine Hyperebene ist eine $n-1$ -dimensionale Fläche, die einen n -dimensionalen Raum in zwei Hälften teilt. Im

einander übergeführt werden können. In Beispiel Abbildung 3.12 ist die Permutations-Hyperebene durch die gestrichelt dargestellte ansteigende diagonale Linie visualisiert. Als Beispiel für eine Permutation sind die Punkte 6-3 ($d-f$) und 3-6 ($f-d$) markiert.

- Die *Transpositionsäquivalenz* verbindet alle Punkte in \mathbb{R}^n , deren orthogonale Projektion¹⁸ auf eine bestimmte Hyperebene H identisch ist. Die Hyperebene H wird dabei durch die Akkorde gebildet, deren Tonnummern in der Summe 12 bzw. 0 (mod 12) ergeben. In Abbildung 3.12 ist diese Hyperebene in Form der gepunktet dargestellten abfallenden Diagonalen eingezeichnet.
- Die *musikalische Invertierung* wird durch geometrische Invertierung durch den Koordinatenursprung dargestellt. *Kardinalitätsäquivalenz* verbindet Punkte von Räumen mit unterschiedlichen Dimensionen

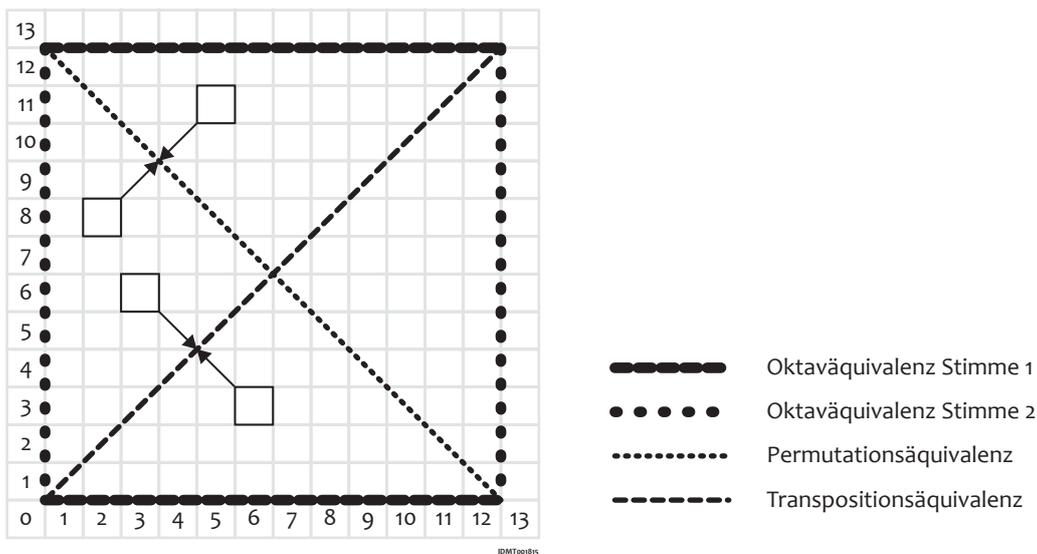


Abbildung 3.12 – Darstellung verschiedener Äquivalenzrelationen im Raum \mathbb{Z}^2 . Callenders, Quinns und Tymoczkos Idee besteht darin, den Raum derartig zu transformieren, dass äquivalente Punkte zusammenfallen.

Falle einer Linie ist dies ein Punkt, im Fall einer Fläche eine Linie, im Fall eines dreidimensionalen Raumes eine Fläche und im Fall eines n-dimensionalen Raumes eine n-1-dimensionale Fläche.

¹⁸ Legt man in einen n-dimensionalen Raum eine n-1 dimensionale Hyperebene, so kann man alle Punkte orthogonal auf diese Ebene projizieren. Im Beispiel der zweidimensionalen Fläche kann man eine eindimensionale Linie in diesen Raum legen und alle Punkte auf diese Linie projizieren.

In Abbildung 3.13a ist der Stimmführungsraum unter Annahme von Transpositionsäquivalenz zu sehen. Dies bedeutet, dass in diesem Raum Akkorde nur noch hinsichtlich ihrer inneren Intervallstruktur verglichen werden. Die Akkorde *c-e-g* und *d-fis-a* werden dadurch in Abbildung 3.13a durch den gleichen Punkt $(0, 4, 7)$ repräsentiert. Durch Hinzunahme weiterer Äquivalenzrelationen werden weitere Punkte des Raumes als äquivalent zusammengefasst, die Anzahl der im Raum befindlichen Punkte sinkt dadurch. Dies ist in Abbildung 3.13b zu sehen: Hier kommt noch die Äquivalenzrelation „Invertierungsäquivalenz“ hinzu, was zu einem dreieckigen Raum führt. Durch die Invertierungsäquivalenz werden z.B. Dur- und Molldreiklänge als äquivalent betrachtet. In Abbildung 3.13b sind beide Akkordtypen durch den Punkt $0, 3, 7$ repräsentiert.

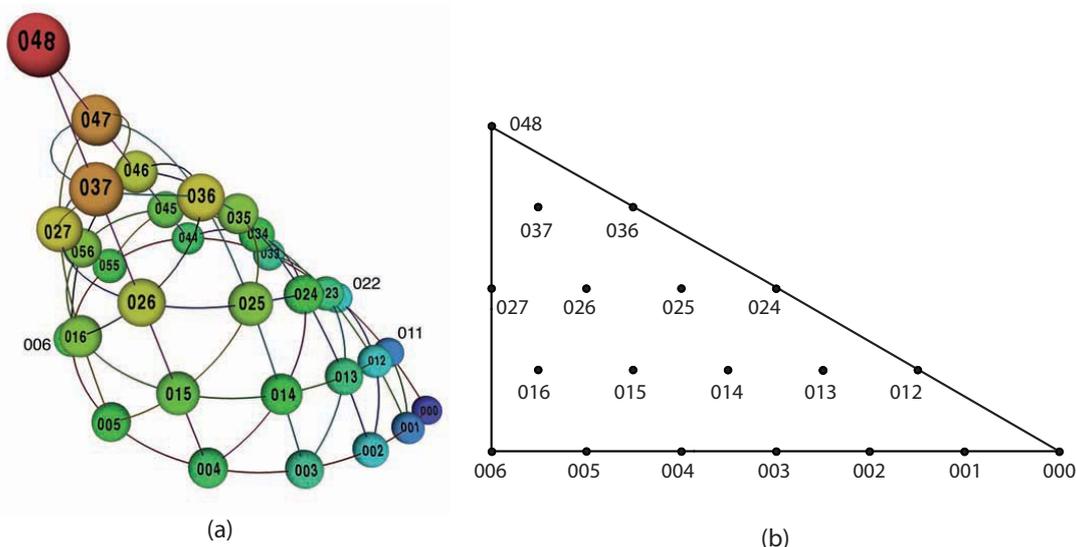


Abbildung 3.13 – (a) Die Punkte repräsentieren äquivalente Dreitonkombinationen unter Annahme von Transpositionsäquivalenz. (b) Die Punkte repräsentieren äquivalente Akkorde unter Annahme von Transpositions- und Invertierungsäquivalenz [19]. Die Zahlen $0, 1, 2, \dots$ repräsentieren die Tonigkeiten *c, cis, d, \dots*

Bewertung:

Nach Aussage der Autoren sind die vorgestellten Tonräume geeignet für die Modellierung musikalischer Ähnlichkeit und zur Analyse einer großen Weite zeitgenössischer Musik. Hinsichtlich der Wahrnehmung von atonalen Akkorden haben Tuire Kuusi [87], [88], Arthur G. Samplaski [138] und Cheryl L. Bruner [16] herausgefunden, dass hier die wahrgenommene Ähnlichkeit vor allem durch die Anzahl der Stimmen, die Menge gemeinsamer Töne zweier Akkorde, die Ähnlichkeit von atonalen Akkorden zu traditionellen tonalen Akkorden sowie deren Oktavlage, Akkordbreite und Dissonanz bedingt

ist. Inwiefern solche perzeptuellen Ähnlichkeiten durch Callenders, Quinns und Tymoczkos Modell repräsentiert werden, ist zu untersuchen. Weiterhin behaupten die Autoren, dass die durch die Räume wiedergegebenen Äquivalenzrelationen zentraler Bestandteil der Musik seit dem 17. Jahrhundert sein. Das Modell übertrage musiktheoretische Begriffe in eine präzise geometrische Sprache und erlaube eine große Menge mathematischer Rückschlüsse. Hier muss jedoch gesagt werden, dass gerade Grundpfeiler der abendländischen Tonalität wie die Quinte und Terz als zentrale Harmonieträger keine Rolle spielen. Funktionelle Aspekte können deshalb in das Modell ebensowenig eingeordnet werden wie Aspekte der diatonischen Stimmführung oder melodische oder harmonische Anziehungs- und Spannungskräfte. Kernproblem von Callenders, Quinns und Tymoczkos Stimmführungsräumen ist die auch schon von Fred Lerdahl [101, S. 47] bemerkte Tatsache, dass das Halbtonintervall und nicht die Quinte als Ausgangspunkt der tonalen Hierarchie gesehen wird¹⁹. Veränderung einzelner Stimmen eines Akkordes um einen Halbton führen musikpsychologisch gesehen in vielen Fällen zu drastischen Änderungen der Akkordwirkung.^{20,21} Zusammenfassend kann gesagt werden, dass die Bedeutung von Callenders, Quinns und Tymoczkos Modell in der geometrischen Visualisierung der speziell für atonale Musik entwickelten Pitch-class Set Theory liegt. Das Modell ist jedoch nicht geeignet, um musiktheoretische und musikpsychologische Eigenschaften der abendländischen Tonalität zu analysieren oder Musikstücke hinsichtlich ihrer wahrgenommenen Wirkung zu vergleichen.

3.10 Dasian-, Guidonian- und Affinitätsräume

In seiner Dissertation beschäftigte sich José António Oliveira Martins [107] mit der Analyse von Musik des 20. Jahrhunderts, darunter von Komponisten wie Béla Bartók, Igor Stravinsky und Darius Milhaud. Ein sehr häufig auftretendes Merkmal dieser Musik ist die nichttonale Aufeinanderfolge oder Überlagerung diatonischer Melodiefragmente

¹⁹ Tymoczko selber schreibt dazu [163]: „In earlier periods, non-diatonic tones were typically understood to derive from diatonic tones: for example, in C major, the pitch class F# might be conceptualized variously as the fifth of B, the leading tone of G, or as an inflection of the more fundamental diatonic pitch class F. By the start of the twentieth century, however, the diatonic scale was increasingly viewed as a selection of seven notes from the more fundamental chromatic collection. No longer dependent on diatonic scale for their function and justification, the chromatic notes had become entities in their own right.“

²⁰ Verändert man einzelne Tonigkeiten eines Dur- oder Mollakkordes um einen Halbton, so ergeben sich völlig andere Klänge.

²¹ Weitere musiktheoretische Widersprüche eines der Unterräume von Callenders, Quinns und Tymoczkos Modell werden auch von Dave Headlam und Matthew Brown [64] bemerkt.

(Diatonic Strata) aus ganz unterschiedlichen Tonarten. Um melodische und harmonische Distanzen solcher Skalensegmente zu messen, schlägt Martins zwei kreisförmige Tonräume vor, nämlich den *Dasian Space* und den *Guidonian Space*. Inspiriert durch das mittelalterliche auf Hexachorden basierte melodische System werden diese Tonräume durch die Aneinanderreihung sogenannter *modularer Einheiten* (Modular Units) gebildet. Eine modulare Einheit besteht dabei im Fall des Dasian Space aus vier und im Fall des Guidonian Space aus drei Ganztönen. Der Abstand zwischen zwei benachbarten modularen Einheiten beträgt einen Halbtonschritt (in Abbildung 3.14 durch Verbindungsclammern zwischen Tonigkeitsnamen gekennzeichnet). Um den harmonischen Verlauf von Musikstücken zu charakterisieren, führt Martins verschiedene Bewegungstypen ein.

- Dazu gehört die *schrittweise Bewegung* (Stepwise Motion), bei der die Melodie ausgehend von einer bestimmten Tonigkeit im Dasian oder Guidonian Space im oder gegen den Uhrzeigersinn fortfährt. Der Melodieverlauf c, d, e, f, g ist z.B. eine schrittweise Bewegung in den in Abbildung 3.14 gezeigten Tonräumen. Die Stepwise Motion ist die einfachste und am häufigsten verwendete melodische Bewegung.
- Der zweite Bewegungstyp wird als *Channeling* bezeichnet und soll an einem Beispiel erläutert werden: In Abbildung 3.14a ist im Dasian Space die Tonigkeit C markiert. Sie befindet sich an vier mit (1), (2), (3) und (4) markierten Stellen im Tonraum. Erklängt nun in einem Musikstück die Tonigkeitsfolge $g, a, b (h), c$, so befindet sich das Musikstück an der in Abbildung 3.14a mit (1) markierten Stelle. Channeling bedeutet nun, dass die Melodie des Stückes nicht an der Stelle (1) weiter fortfährt, sondern an eine der Stellen (2), (3) oder (4) springt. Im Beispiel der Stelle (4) würde die Melodie z.B. mit den Tonigkeiten c, des, f und es weiter fortfahren.
- Der dritte Bewegungstyp wird als *chromatische Inflektion* (*Chromatic Inflection*²²) bezeichnet, was die Erhöhung oder Erniedrigung einer einzelnen Tonigkeiten um einen Halbton bezeichnet. Nehmen wir an, eine Melodie befindet sich z.B. im melodischen Kontext $g, a, b (h), c$ und damit in Abbildung 3.14b an der mit (1) markierten Stelle. Nun wird die Tonigkeit $b (h)$ zum b_b (b) erniedrigt. Die Melodie befindet sich dadurch nun im Bereich g, a, b_b, c , welcher in Abbildung 3.14b mit (2) markiert ist. Der Wechsel vom Kontext (1) auf den Kontext (2) wird als chromatische Inflektion bezeichnet.

²² Als „Inflection“ wird die absichtliche Abweichung von der Norm eines Tones bezeichnet. Ein sehr häufig verwendetes Beispiel ist die Blue Note im Jazz [131].

- Der vierte Bewegungstyp ist die *Mutation*. Als Mutation wird im Kontext der mittelalterlichen Musiktheorie der Wechsel von einem Hexachord auf einen anderen bezeichnet [85]. Im Kontext von Martins Tonräumen bezeichnet der Begriff Mutation eine Bewegung im Guidonian Space um sieben Schritte. Z.B. der in Abbildung 3.14b eingezeichnete Wechsel des Hexachords f, g, a, b, c, d auf den Hexachord g, a, h, c, d, e ist eine Mutation. Eine Mutation führt wahlweise zur chromatischen Inflektion oder zum Channeling von Tonigkeiten. Im vorherigen Beispiel werden die Tonigkeiten g, a und c gechannelt, die Tonigkeiten h und c werden chromatisch inflektiert.

Martins analysiert auf Basis kombinierter Mutationsoperationen verschiedene Werke des 20. Jahrhunderts. Als eine Verallgemeinerung des Dasian und des Guidonian Space schlägt er schließlich noch sogenannte *Affinity Spaces* vor.²³ Das sind Räume, die periodisch zwei Intervalle alternieren.

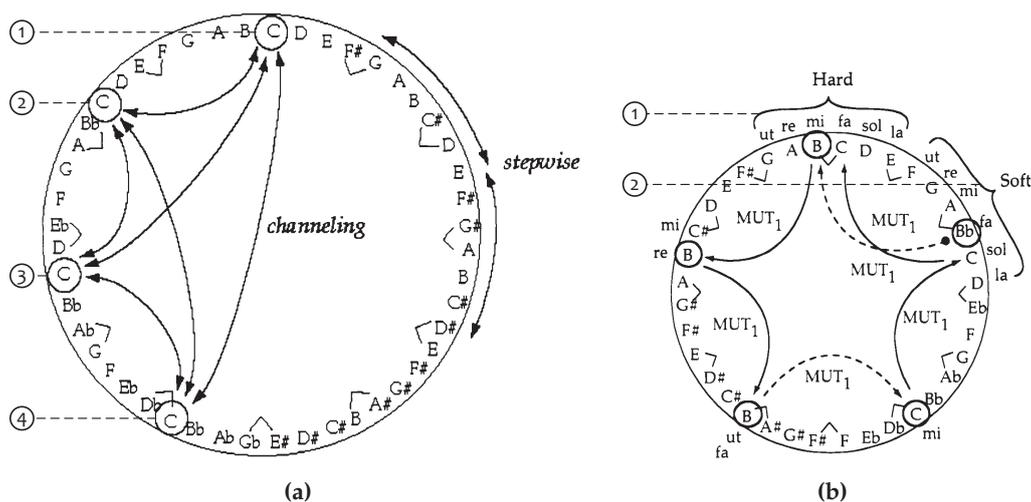


Abbildung 3.14 – (a) Der Dasian Space, (b) Der Guidonian Space [107]

Bewertung:

Die Zerlegung polyphoner Musikstücke in diatonische Melodiefragmente und die Einordnung dieser Fragmente in geometrische Modelle wie den Guidonian und den Dasian Space ist eine sehr interessante Möglichkeit zur melodischen Analyse und Klassifikation sowohl tonaler Musik, von Jazzmusik, als auch von neuer Musik. Eine Klassifikation von Musikstücken hinsichtlich der Häufigkeit der Operationen „Channeling“, „schrittweise

²³ Affinity = Wesensverwandtschaft

Bewegung“ und „chromatische Inflektion“ könnte durchaus zu brauchbaren Ergebnissen führen. Die Klassifizierung tonaler Bewegungen in diese drei Typen ist auch für die Entwicklung tonraumbasierter Musikinstrumente interessant. Hier muss man Möglichkeiten finden, tonale Bewegungen einfach darzustellen. Martins Ansatz ist ein möglicher Ausgangspunkt. Der Bezug von Martins Kreissystemen zur Diatonik und die quintorientierte Anordnung der in den Räumen enthaltenen diatonischen Tonleitern lässt die Möglichkeiten offen, dass die in den Räumen dargestellten geometrischen Verhältnisse auch wahrgenommene Ähnlichkeiten widerspiegeln. An dieser Stelle sind jedoch noch empirische Untersuchungen erforderlich, welche die einzelnen Bewegungen im Tonraum untersuchen.

3.11 Weitere Arbeiten

Moritz Hauptmanns [63] verwendete eine Tonreihe, bei der benachbarte Töne abwechselnd ein Intervall von einer großen oder kleinen Terz besaßen, um den Aufbau von Dur- und Molldreiklängen, deren Funktion innerhalb einer Tonart (Tonika, Oberdominante, Unterdominante) sowie den Zusammenhang zwischen Tonarten zu erläutern. Im Symmetriemodell (Kapitel 4) findet sich Hauptmanns' System in Form des tonartübergreifenden Terzkreises wieder (Abschnitt 4.2.4).

Hendrik Purwins [127] leitete eine toroidale Darstellung des Vial/Weber/Schönbergnetzes durch statistische Analysen musikalischer Daten auf Basis von Ansätzen neuronaler Netzwerke bzw. selbstorganisierender Karten her. Ebenfalls auf Basis von selbstorganisierenden Karten arbeitet Toiviainen [161] Ansatz zur Visualisierung der dynamischen Tonartenentwicklung in einem Musikstück.

Craig Stuart Sapp [140] schlägt ein System zur Visualisierung der in einem Musikstück verwendeten Tonarten vor. Der Algorithmus berücksichtigt die Tatsache, dass die Tonart nicht für ein gesamtes Musikstück analysiert werden kann, sondern dass es auch möglich ist, ein Musikstück in Segmente zu zerlegen, welche wiederum eine eigene Tonart besitzen. Die gefundenen Segmente können wieder in Subsegmente zerlegt werden, welche ebenfalls eine Tonart besitzen. Auf einer ersten Achse eines zweidimensionalen Koordinatensystems repräsentiert Sapp deshalb die Zeit. Auf einer zweiten Achse wird das untersuchte Musikstück jeweils in ein, zwei, drei, vier und mehr Segmente zerlegt. Die in den jeweiligen Segmenten vorliegende Tonart wird farblich gekennzeichnet. Das resultierende Diagramm ermöglicht so, die harmonische Struktur eines Musikstückes zu analysieren oder auch die Ausgabe verschiedener Tonarterkennungsalgorithmen zu vergleichen.

Andere Arbeiten zu mathematisch-geometrischen Tonalitätsansätzen erfolgten unter anderem durch Roger Shepard [152], Gottfried Steyer [154], Werner Pöhlert [128], Guerino Mazzola [111], [109] und Thomas Noll [120], [119].

3.12 Tonraumbasierte Analyse und Synthese von Musik

Tonräume werden zum einen bei der Analyse und Klassifikation von Musiksignalen eingesetzt. Dies ist Gegenstand ab Kapitel 7. Zum anderen können Tonräume eingesetzt werden, um Audiosignale zu erzeugen. Im Folgenden sollen Arbeiten genannt werden, die geometrische Tonraumsysteme zur Erzeugung von Musik verwenden.

1987 bis 1994 entwickelte Guerino Mazzola zusammen mit der Fraunhofer Arbeitsgruppe für grafische Datenverarbeitung in Darmstadt die Musikkompositionssoftware „Presto“ (Music Designer Z71) [109], [110]. Die Software bietet die Möglichkeit, sogenannte Motive bzw. lokale Kompositionen in einem Notenrollen-Editor – „Operating Score“ genannt – zu definieren (Abbildung 3.15). Weiterhin können verschiedene Symmetrietranslationen definiert und auf das zuvor definierte Motiv angewendet werden. Das resultierende Tonfragment wird als „Ornament“ bezeichnet. Teile des Ornamentes können anschließend in der „Score“ auf eine bestimmte metrische oder harmonische Struktur quantisiert und zeitlich arrangiert werden.

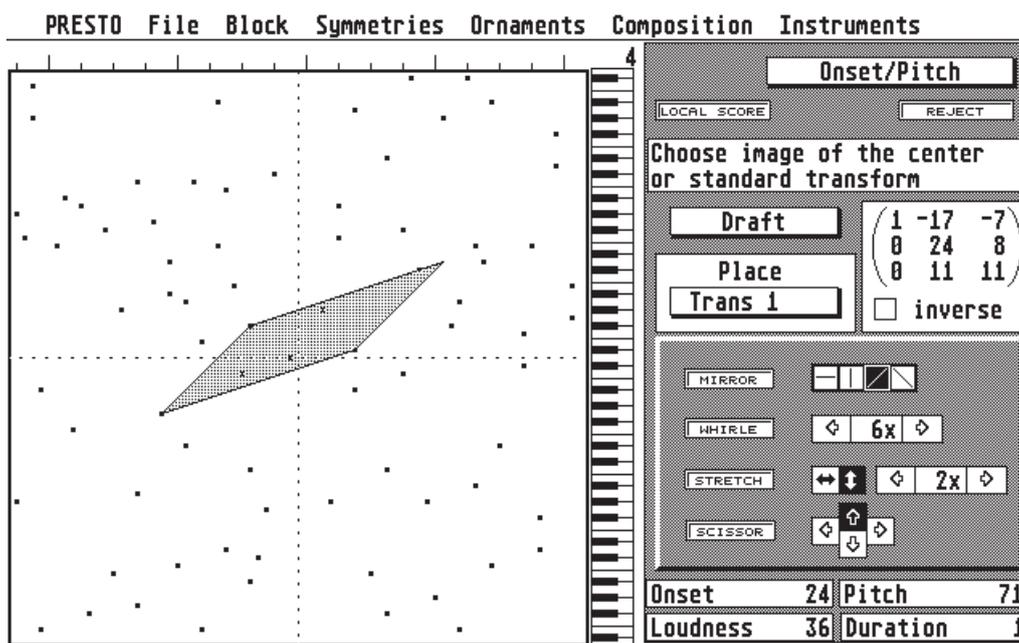


Abbildung 3.15 – Der Music Designers Z71 [6]

Die aktuelle Weiterentwicklung von Presto ist der sogenannte „Rubato Composer“²⁴ von Gérard Milmeister [113], [114]. Die Software stellt eine an grafisch-objektorientierte Programmiersprachen angelehnte Softwareoberfläche bereit, mit deren Hilfe Module – Ru-

²⁴ www.rubato.org

betten genannt – zu sogenannten Netzwerken verbunden werden. Rubetten wiederum arbeiten auf Basis von *Dentatoren*, das sind Instanzen (Objekte) von sogenannten *Formen* (Forms), welche Klassen in objektorientierten Programmiersprachen entsprechen. Mit dem Rubato Composer werden bereits eine Reihe von Rubetten bereitgestellt, darunter auch die Rubette *OrnaMagic* von Florian Thalmann [160], welche die bereits im Zusammenhang mit der Software Presto beschriebene Ornamentfunktionalität als Rubato Modul anbietet.

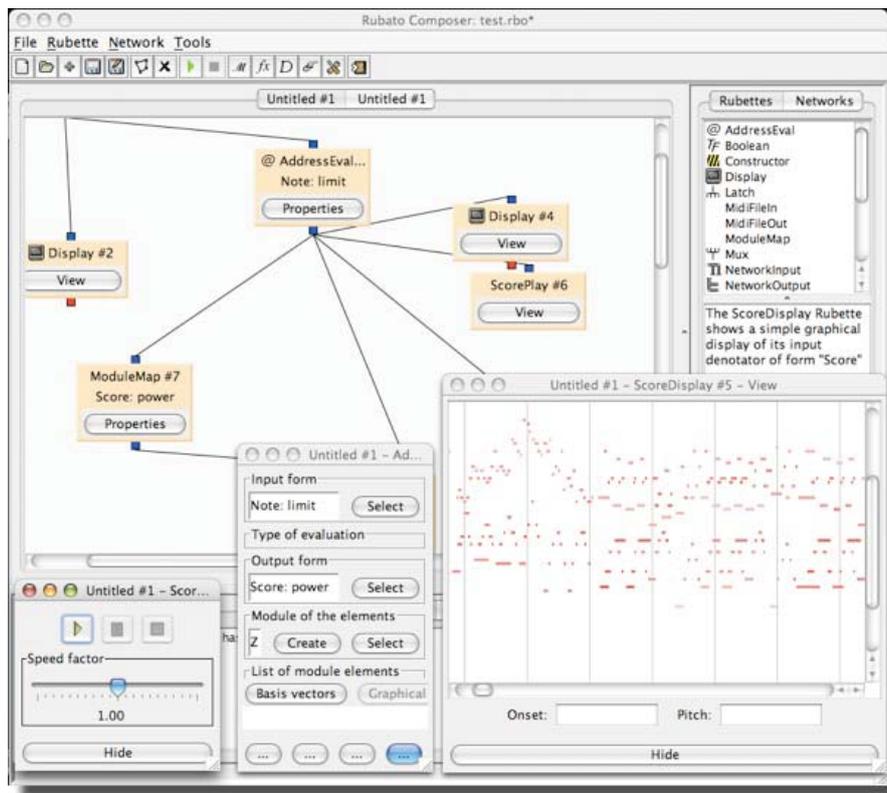


Abbildung 3.16 – Der Rubato Composer [6]

Ein innovatives kommerzielles Musikinstrument, das sich von bisherigen klaviaturbasiereten Ansätzen deutlich unterscheidet, ist das Yamaha TENORI-ON [118], [5]. Kernkomponente des Instrumentes ist eine 16x16 LED-Tastenmatrix, auf welcher der Spieler grafische Muster „zeichnen“ kann (Abbildung 3.17). Je nach gewähltem Modus wird das erstellte zweidimensionale Muster unterschiedlich interpretiert. Im einfachsten Fall durchläuft ein Zeitmarker die Matrix von links nach rechts. Bei Zusammentreffen des Zeitmarkers mit dem erstellten Muster werden die dem „berührten“ Musterabschnitt entsprechenden Töne ausgegeben. Komplexere Modi ersetzen den Zeitmarker durch andere Paradigmen, z.B. durch einen sich frei durch die Matrix bewegenden Ball.



Abbildung 3.17 – Das TENORI-ON [5]

Ebenfalls geometrische Ansätze wurden bereits mehrfach in musiktheoretischen Lernhilfen umgesetzt. Kombiniert mit einer Klangerzeugung können derartige Werkzeuge zu neuartigen Musikinstrumenten weiterentwickelt werden (siehe Kapitel 6). Bei solchen Werkzeugen handelt es sich meist um Schablonen, Scheiben oder andere Objekte, welche mechanisch miteinander verbunden werden und gegeneinander verschiebbar oder verdrehbar sind. Auf den Schablonen werden musiktheoretische Zusammenhänge dargestellt. Beispiele für derartige Werkzeuge sind [103], [68], [46], [99], [48], [91], [136], [146], [147] und [143].

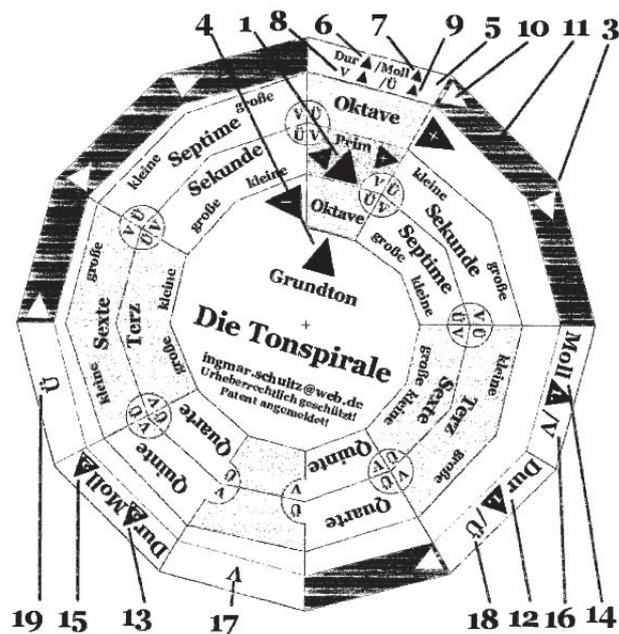


Abbildung 3.18 – Beispiel einer Lernschablone [143]

Ein an das Eulernetz (Abschnitt 3.4) angelehntes Musikinstrument wird in [13] vorgeschlagen. Das Musikinstrument ist durch eine Anordnung sechseckiger Tasten – wie in Abbildung 3.19 – gezeigt gekennzeichnet. Durch dieses bienenwabeförmige Tastenlayout bilden sich Tastenzeilen in jeweils drei Richtungen. Die Töne sind den Tasten so zugeordnet, dass die erste der benannten Tastenzeilen eine Tonfolge in Quinten bildet, die zweite eine Tonfolge in kleinen Terzen und die dritte eine Tonfolge in großen Terzen. Durch diese Anordnung können die in Abschnitt 3.4 beschriebenen Eigenschaften des Eulernetzes spielerisch erfahren werden. D.h. durch Drücken in das "Kreuz" aus drei angrenzenden Tasten kann ein Dur- oder Mollakkord gespielt werden. Durch spielen benachbarter "Kreuze" können die entsprechend quint- und terzverwandten Akkorde leicht gefunden werden. Das Instrument ist inzwischen unter dem Namen AXiS-49 auf dem Markt.²⁵

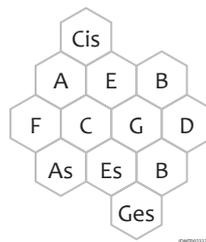


Abbildung 3.19 – Tastenlayout des Eulernetz basierten Musikinstrumentes (nach [13])

²⁵ http://www.c-thru-music.com/cgi/?page=prod_axis-49

Kapitel 4

Das Symmetriemodell

Nachdem der Stand der Technik auf dem Gebiet geometrischer Tonraummodelle vorgestellt wurde, wird nun mit dem Symmetriemodell ein eigenes neuartiges Tonraummodell eingeführt. Das Symmetriemodell beschreibt wichtige Eigenschaften der abendländischen Tonalität in einer äußerst kompakten Art und Weise. Dazu gehören die unterschiedlichen Ebenen der Harmoniewahrnehmung, die Modellierung der kontext- bzw. tonartbezogenen Tonalitätswahrnehmung, funktionstheoretische Aspekte, die Gruppierung von Tönen zu Akkorden und Tonleitern sowie eine Vielzahl weiterer aus der Musiktheorie schon sehr lange bekannter Gesetzmäßigkeiten. Das Symmetriemodell besitzt deshalb großes Potential, in praktischen Anwendungen wie der Tonart- und Akkorderkennung, der Berechnung harmonischer Ähnlichkeit oder auch der Entwicklung neuartiger Musikinstrumente oder elektronischer Musikspiele zum Einsatz zu kommen. Dies ist jedoch Gegenstand späterer Kapitel. Das vorliegende Kapitel ist folgendermaßen gegliedert: Zunächst werden in Abschnitt 4.1 verschiedene Grundbegriffe sowie mathematische Grundlagen zur allgemeinen Beschreibung kreisförmiger Tonräume bereitgestellt. Anschließend werden in Abschnitt 4.2 die verschiedenen Teilräume des Symmetriemodells hergeleitet. Die spiralförmige Darstellung des Symmetriemodells sowie verschiedene Erweiterungen des Modells werden in Abschnitt 4.3 behandelt.

4.1 Allgemeine Beschreibung kreisförmiger Tonräume

Zunächst sollen Begrifflichkeiten zur allgemeinen Beschreibung kreisförmiger Tonräume eingeführt werden. Auch wenn das Symmetriemodell erst im nachfolgenden Abschnitt 4.2 hergeleitet wird, sei jetzt schon vorweggenommen, dass einige der nachfolgend beschriebenen Konventionen im Hinblick auf eine symmetrische Verteilung von Tönen um einen sogenannten Symmetrieton ausgerichtet sind.

4.1.1 Verwendetes Koordinatensystem

Für die Darstellung von kreisförmigen Tonräumen verwenden wir ein Polarkoordinatensystem, das wie folgt definiert ist: Der Winkel 0° liegt auf der Y-Achse und steigt im positiven Uhrzeigersinn an. Motiviert ist diese Konvention durch den Quintenzirkel, wo die vorzeichenlose Tonart C-Dur oben dargestellt ist. Tonarten mit negativem Vorzeichen (b) befinden sich gegen den Uhrzeigersinn von C-Dur aus, Tonarten mit positivem Vorzeichen (#) im Uhrzeigersinn.

4.1.2 Hilfstonigkeiten und Realtonigkeiten

Die Grundlage für alle nachfolgend beschriebenen Tonsysteme sind die sogenannten *Hilfstonigkeiten* ξ . Die Hilfstonigkeiten werden wie in Abbildung 4.1 durch graue Kreise repräsentiert. Zwei benachbarte Hilfstonigkeiten haben immer einen Abstand von einem Halbton. Dadurch definieren sie ein Halbtonraster und stellen somit einen Bezug zu physikalischen Intervallverhältnissen her. Den nachfolgenden Ausführungen wird eine temperierte Stimmung zugrunde gelegt, was in Form gleichabständiger Hilfstonigkeiten visualisiert wird. Setzte man eine andere Stimmung voraus – wie zum Beispiel die reine Stimmung oder die pythagoreische Stimmung – so würde der Abstand der Hilfstonigkeiten je nach Kontext variieren. Die Menge aller Hilfstonigkeiten M_ξ ist folgendermaßen definiert:

$$M_\xi := \mathbb{Z} \quad (4.1)$$

Neben den Hilfstonigkeiten ξ gibt es die sogenannten *Realtonigkeiten* ρ , welche die tatsächlich in die musikalische Analyse einbezogenen Tonigkeiten repräsentieren. Die Realtonigkeiten ρ werden durch schwarze Kreise dargestellt (Abbildung 4.1). Die Menge an Realtonigkeiten M_ρ ist von Tonraum zu Tonraum verschieden und ist eine Untermenge der Hilfstonigkeiten M_ξ :

$$M_\rho \subseteq M_\xi \quad (4.2)$$

Jede Realtonigkeit ρ ist somit auch gleichzeitig eine Hilfstonigkeit ξ . In Abbildung 4.1 wird dies durch die graue Hinterlegung visualisiert.

4.1.3 Halbtonauflösung

Die Anzahl der Hilfstonigkeiten, aus denen ein kreisförmiger Tonraum besteht, wird als Halbtonauflösung g bezeichnet. Die Halbtonauflösung g des in Abbildung 4.1 dargestellten Tonraumes beträgt 12. Es gilt:

$$g \in \mathbb{N}, g \bmod 2 = 0 \quad (4.3)$$

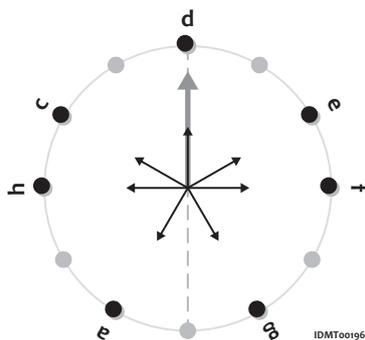


Abbildung 4.1 – Beispiel eines kreisförmigen Tonraumes

4.1.4 Die Symmetrieachse und Symmetrietonigkeit

In Anlehnung an David Gatzsches Arbeit ([47], Abschnitt 3.8) richten wir das Koordinatensystem des kreisförmigen Tonraumes so aus, dass die *Symmetrietonigkeit* eines diatonischen Tonsets im Kreis entweder *oben* oder *unten* dargestellt wird. In Beispiel von Abbildung 4.1 wird dies für das diatonische Tonset C-Dur gezeigt. Das Koordinatensystem ist so ausgerichtet, dass sich dabei die Symmetrietonigkeit *d* oben befindet. Bei Wahl eines anderen diatonischen Tonsets wäre dies eine andere Tonigkeit. Als *Symmetrieachse* wird die vertikale durch die Symmetrietonigkeit verlaufende Achse bezeichnet. Sie wird in Form einer grau gestrichelten Linie dargestellt (Abbildung 4.1). Es ist weiterhin wichtig, anzumerken, dass die Symmetrietonigkeit hauptsächlich *eine Hilfstonigkeit* ist und nur im Falle der später eingeführten Tonräume *F* und *FR* mit einer Realtonigkeit zusammenfällt.

4.1.5 Grund- und erweiterter Index

Um jede Hilfstonigkeit ξ eindeutig zu identifizieren, wird der *erweiterte Index* $n \in \mathbb{Z}$ eingeführt. Der Index $n = 0$ repräsentiert den Symmetrieton der vorzeichenlosen Tonart C-Dur (Abschnitt 4.2.3). Hilfstonigkeiten mit Indizes $n > 0$ befinden sich in Uhrzeigerrichtung von $n = 0$, Hilfstonigkeiten mit Indizes $n < 0$ befinden sich gegen Uhrzeigerrichtung von $n = 0$. Der erweiterte Index n erlaubt es, die auf einem Kreis angeordneten Tonigkeiten durchzuzählen, ohne Rücksicht auf mehrfache Kreisdurchläufe zu nehmen. Dies bedeutet auch, dass eine in einem Kreis dargestellte Hilfstonigkeit ξ unendlich viele erweiterte Indizes $n + k * g, k \in \mathbb{Z}$ besitzt. Um auch eine eineindeutige Zuordnung von Tonigkeiten und Indizes zu haben, wird der *Grundindex* m eingeführt. Unter Kenntnis der Halbtonauflösung g des Kreises gilt:

$$\frac{-g}{2} \leq m < \frac{g}{2} \quad m \in \mathbb{Z} \quad (4.4)$$

Für den gemeinsamen Definitionsbereich sind Grundindex m und erweiterter Index n identisch. D.h. die Zuordnung des Grundindex m zu Hilfstonigkeiten des Tonraumes entspricht der in Abschnitt 4.1.5 gegebenen.

4.1.6 Vektorrepräsentation von Tonigkeiten

In einem kreisförmigen Tonraum ist es möglich, eine Realtonigkeit ρ in Form eines komplexen Vektors $\vec{\rho}$ zu repräsentieren (Abbildung 4.1, schwarze dünne Pfeile). Der Winkel α des Vektors wird durch die Position der Tonigkeit im Tonraum bestimmt, während die Energie¹ ε der Tonigkeit auf den Betrag des Vektors abgebildet wird. Unter Einbezug der Halbtonauflösung g des Kreises gilt:

$$\vec{\rho} = \varepsilon * e^{j*\alpha}, \alpha = 2\pi * \frac{n}{g} \quad \alpha \in \mathbb{R} \quad (4.5)$$

Addiert man die komplexen Vektoren aller Realtonigkeiten ρ , so erhält man einen Summenvektor $\vec{\Psi}$ (Abbildung 4.1, fetter grauer Pfeil):

$$\vec{\Psi} = \sum \vec{\rho}_m \quad m \in M_\rho \quad (4.6)$$

Mit Hilfe eines solchen Summenvektors ist es möglich, musikalische Eigenschaften eines Musikstückes oder -abschnittes stark zusammenzufassen und damit niedrigdimensional zu repräsentieren. Die ausführliche Diskussion des Summenvektors vieler wichtiger Tonkombinationen ist Gegenstand von Kapitel 5.

4.1.7 Veränderte Indexzuordnung zu chromatischen Tonigkeiten

Einzelne Tonigkeiten des chromatischen Tonsets (Abschnitt 2.6.2) werden im Kontext des Symmetriemodells über einen von den Standardindizes (Tabelle 2.1) abweichenden Index identifiziert. Die Identifikation einer chromatischen Tonigkeit erfolgt über die Indizes m_C oder n_C (siehe Abschnitt 4.1.5). Im Gegensatz zum Stand der Technik (Tabelle 2.1) wird nicht die Tonigkeit c mit dem Grundindex $m_C = 0$ versehen, sondern die Symmetrietonigkeit des diatonischen Tonsets C-Dur, das ist die Tonigkeit d (Abschnitt 4.1.4). Dadurch ergeben sich verschiedene Vereinfachungen, z.B. bei der Einordnung von Tonigkeiten in das Symmetriemodell (Abschnitt B.1):

¹ Die Energie kann eine physikalische Größe, aber auch eine gehörbewertete psychoakustische Größe sein.

<i>Tonigkeit</i>	...	a	<i>ais</i>	h	c	<i>cis</i>	d	<i>dis</i>	e	f	<i>fis</i>	...
n_C	...	-5	-4	-3	-2	-1	0	1	2	3	4	...

Tabelle 4.1 – Zuordnung von erweiterten Indizes n_C zu Tonigkeiten.

4.1.8 Zusammenfassung der Symbole

g	Halbtonauflösung des Kreises
ξ	Hilfstonigkeit
ρ	Realtonigkeit
m	Grundindex einer Hilfstonigkeit
n	Erweiterter Index einer Hilfstonigkeit
ρ_m	Realtonigkeit mit Grundindex m
ρ_n	Realtonigkeit mit erweitertem Index n
M_ρ	Die Menge der Realtonigkeiten eines Tonraumes
M_ξ	Die Menge der Hilfstöne eines Tonraumes
α_{ρ_m}	Winkel der Realtonigkeit mit dem Grundindex m
$\vec{\Psi}$	Summenvektor im kreisförmigen Tonraum

Tabelle 4.2 – Zusammenfassung aller Begriffe und mathematischen Symbole, um kreisförmige Tonräume zu beschreiben

4.2 Die Teilräume des Symmetriemodells

Das Symmetriemodell besteht aus einer Menge von Teilräumen, welche im Folgenden eingeführt werden. Durch die Beachtung des Symmetrietones wird es gelingen, einige zum Teil unabhängig existierende Tonräume deutlich einfacher als bisher zu beschreiben und in einem einheitlichen mathematisch-geometrischen Modell zu vereinen. Weiterhin wird es möglich sein, Aspekte der abendländischen Tonalität zu modellieren, die mit geometrischen Modellen bisher nur sehr umständlich oder gar nicht beschrieben werden konnten. Zu diesen Aspekten gehören z.B. die in Abschnitt 5.3.6 diskutierten Alterationen oder die tonale Stabilität von Tönen in einem gegebenen Kontext (Abschnitt 5.5.4). Im vorliegenden Abschnitt geht es jedoch zunächst nur um die rein mathematisch-geometrische Beschreibung des Symmetriemodells.

4.2.1 Die Symmetriemodell-Generatorformel

Der Ausgangspunkt des Symmetriemodells ist die sogenannte Symmetriemodell-Generatorformel. Diese beschreibt die vier wichtigsten tonalen Hierarchieebenen der westlichen Musik wie folgt:

$$H = 7\lambda \pm 12 \pm 6 \pm 3 \qquad \lambda \in \mathbb{Z} \qquad (4.7)$$

Die Reihe 7λ beschreibt die *Quintreihe* als Basis des abendländischen Tonalitätssystems. Addiert man zu jedem Element der Quintreihe die Konstante ± 12 , so entsteht die in Abschnitt 4.2.4 beschriebene *Terzreihe* $7\lambda \pm 12$, welche alle Dur- und Molldreiklänge enthält. Die wiederholte Addition einer Konstanten ± 3 zu jedem Element der Terzreihe führt zur sogenannten *diatonischen Tonleiterreihe* $7\lambda \pm 12 \pm 6$. Diese enthält alle diatonischen Tonleitern. Eine weitere Transformation der diatonischen Tonleiterreihe durch Addition einer Konstanten ± 3 führt schließlich zur *chromatischen Reihe* $7\lambda \pm 12 \pm 6 \pm 3$. Die wichtigste Eigenschaft der Symmetriemodell-Generatorformel besteht darin, dass sie die Entstehung der Hauptebenen des abendländischen Tonsystems (Quint-, Terz-, diatonische Tonleiter- und chromatische Ebene) nicht als willkürlich beschreibt, sondern als aus einer Quintreihe heraus entstehend. Interessant ist auch die Rekursivität der Transformation, welche in Form der Konstanten ± 12 , ± 6 und ± 3 hervortritt. Es folgt eine ausführliche Beschreibung der zuvor genannten Ebenen sowie die Herleitung verschiedener tonartbezogener und tonartübergreifender Tonräume.

4.2.2 Der tonartübergreifende Quintenzirkel F

Der erste Tonraum des Symmetriemodells ist der tonartübergreifende Quintenzirkel F , wobei das F für den englischen Term „Fifths“ steht. Dieser entsteht aus der ersten durch die Symmetriemodell-Generatorformel beschriebene Ebene, der Quintreihe 7λ (Abbildung 4.3). Betrachtet man die Quintreihe als eine Menge von Real- und Hilfstonigkeiten (Abschnitt 4.1.2), so wird die Menge der Hilfstonigkeiten durch die Variable λ beschrieben. In der Quintreihe ist jede siebente Hilfstonigkeit (7λ) eine Realtonigkeit, dadurch ergibt sich die Menge der Realtonigkeiten M_{ρ_F} wie folgt:

$$M_{\rho_F} := \{n_F, n_F = 7\lambda, \lambda \in \mathbb{Z}\} \quad (4.8)$$

Legt man eine temperierte Stimmung bzw. enharmonische Äquivalenz zugrunde, so wiederholt sich die Quintreihe alle 84 Halbtöne. Es ist deshalb möglich, die Enden eines aus 84 Halbtönen bestehenden Ausschnittes aus der Quintreihe zu einem Kreis zusammenzuschließen. Der resultierende kreisförmige Tonraum wird als *tonartübergreifender Quintenzirkel F* bezeichnet. Die Halbtonauflösung g_F des Kreises (Abschnitt 4.1.3) beträgt $g_F = 84$ Halbtöne. Der komplexe Vektor einer Realtonigkeit $n_F \in M_{\rho_F}$ im F mit der Energie ε ergibt sich damit entsprechend Gleichung 4.5 wie folgt:

$$\vec{\rho}_F = \varepsilon * e^{j * 2\pi * \frac{n_F}{84}} \quad (4.9)$$

Gemäß Abschnitt 4.1.4 und 4.1.7 stellen wir die Tonigkeit mit dem Index $n_F = 0$, das ist die Symmetrietonigkeit der Tonart C-Dur, im Kreis oben dar. Dadurch ergibt sich die in Abbildung 4.2 gezeigte Zuordnung von Realtonigkeitsindizes und Tonigkeitsnamen.

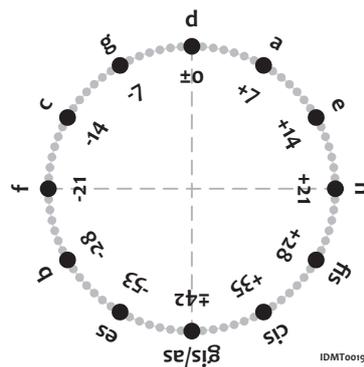


Abbildung 4.2 – Der tonartübergreifende Quintenzirkel F

4.2.3 Der tonartbezogene Quintenzirkel FR

Entnimmt man aus der Quintreihe einen um eine Realtonigkeit symmetrisch verteilten Bereich von 49 Halbtönen, so erhält man eine Tonigkeitsreihe, welche sieben Realtonig-

keiten enthält, die wiederum ein diatonisches Tonset bilden (Abbildung 4.3, gestrichelter Rahmen). Die mittlere dieser sieben Tonigkeiten ist dabei die Symmetrietonigkeit des gewählten diatonischen Tonsets n_k und besitzt den Index $n = 7n_k$.

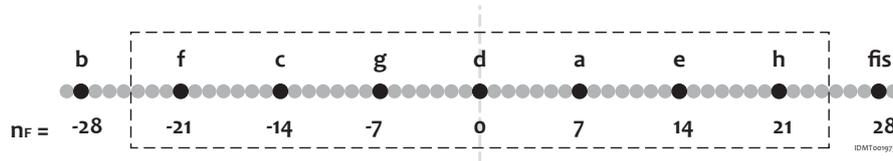


Abbildung 4.3 – Extraktion einer tonartbezogenen Quintreihe aus der Quintebene

Anfang und Ende der entnommenen Reihe sind wie beim tonartübergreifenden Quintenzirkel (enharmonisch) identisch und können deshalb ebenfalls zu einem Kreis geschlossen werden. Dieser wird als *tonartbezogener Quintenzirkel FR* bezeichnet. Das „R“ steht für den englischen Term „key related“. Entsprechend Abschnitt 4.1.4 wird die Symmetrietonigkeit des gewählten diatonischen Tonsets an der Stelle $\alpha = 0$, d.h. oben dargestellt (Abbildung 4.4).

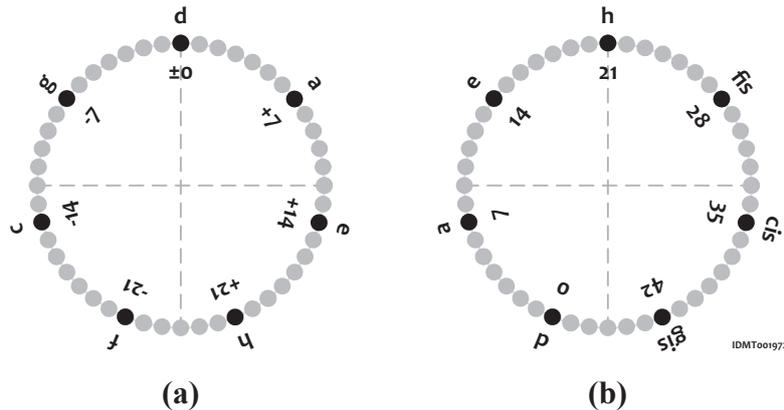


Abbildung 4.4 – Der tonartbezogene Quintenzirkel *FR* für das diatonische Tonset (a) C-Dur und (b) A-Dur

Die Menge der im *FR* abgebildeten Realtonigkeiten $M_{\rho_{FR}}$ ist vom gewählten diatonischen Tonset n_k abhängig und ergibt sich wie folgt:

$$M_{\rho_{FR}} := \{n_{FR}, n_{FR} = 7\lambda + 7n_k, -3 \leq \lambda \leq +3\} \quad (4.10)$$

Der erweiterte Index $n_{FR} = 7n_k$ repräsentiert die Symmetrietonigkeit des ausgewählten diatonischen Tonsets. Die Halbtonauflösung g_{FR} des Kreises beträgt $g_{FR} = 48$ Halbtöne. Der komplexe Vektor einer Realtonigkeit $n_{FR} \in M_{\rho_{FR}}$ im *FR* mit der Energie ε ergibt sich damit entsprechend Gleichung 4.5 wie folgt:

$$\vec{\rho}_{FR} = \varepsilon * e^{j*2\pi*\frac{n_{FR}-7n_k}{48}} \quad (4.11)$$

Die Konstante $-7n_k$ sorgt dafür, dass die Symmetrietonigkeit des gewählten diatonischen Tonsets an der Stelle $\alpha = 0$ dargestellt wird.

4.2.4 Der tonartübergreifende Terzkreis T

Der tonartübergreifende Terzkreis T ist die kreisförmige Repräsentation der bereits im Zusammenhang mit der Symmetriemodell-Generatorformel Abschnitt 4.2.1 angesprochenen Terzreihe. Letztere entsteht – wie in Abbildung 4.5 gezeigt – durch Addition einer Konstanten ± 12 zu jeder Realtonigkeit der Quintreihe 7λ . Die Realtonigkeiten der Quintreihe ergeben sich damit wie folgt:

$$M_{\rho_T} := M_{\rho_F} \pm 12 \tag{4.12}$$

Die Addition der Konstanten ± 12 zu jedem Element der tonartübergreifenden Quintreihe F kann folgendermaßen interpretiert werden: Jede Realtonigkeit der Quintreihe F wird „geteilt“ und eine Oktave (zwölf Halbtöne) nach rechts (+12) und eine Oktave nach links (-12) repliziert (Abbildung 4.5). Dadurch entsteht eine neue Tonreihe, die aus abwechselnd großen und kleinen Terzen besteht und als Terzreihe bezeichnet wird. Die Stärke der Terzreihe besteht darin, dass drei in Benachbarung befindliche Tonigkeiten einen Dur- oder einen Molldreiklang ergeben.

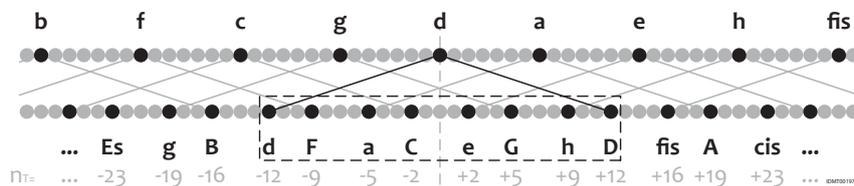


Abbildung 4.5 – Entstehung der Terzreihe aus der Quintreihe

Analog zum tonartübergreifenden Quintenzirkel F kann aus der Terzreihe ein Bereich von 84 Halbtönen (12 Quinten) entnommen und zu einem Kreis geschlossen werden. Dieser wird als *tonartübergreifender Terzkreis T* bezeichnet und ist in Abbildung 4.6 zu sehen. Die Berechnung der komplexen Tonigkeitsvektoren erfolgt wie im F.

4.2.5 Der tonartbezogene Terzkreis TR

Entnimmt man aus der tonartübergreifenden Terzreihe einen um die Symmetrietonigkeit eines diatonischen Tonsets ($n_T = 7n_k$) symmetrisch verteilten Bereich von $24+1$ Halbtönen (Abbildung 4.5, gestrichelter Rahmen), so erhält man eine tonartbezogene Terzreihe. Anfang und Ende dieser Reihe werden durch die gleiche Tonigkeit gebildet und können deshalb wieder zu einem Kreis geschlossen werden. Dieser wird als *tonartbezogener*

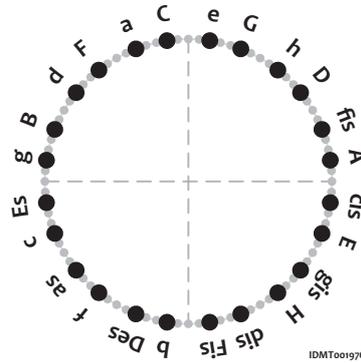


Abbildung 4.6 – Der tonartübergreifende Terzkreis T

Terzkreis TR bezeichnet und ist in Abbildung 4.7 dargestellt. Die Menge der Realtöne im TR sowie deren komplexe Vektorrepräsentation ergeben sich unter Berücksichtigung der Halbtonauflösung $g_{TR} = 24$ analog zum FR (Gleichung 4.11 und Gleichung 4.10).

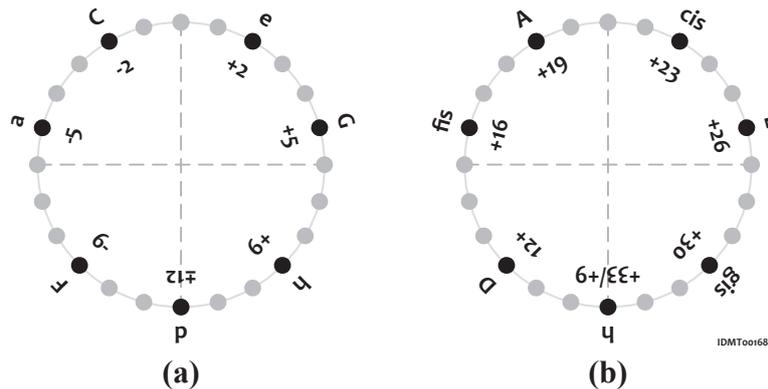


Abbildung 4.7 – Der tonartbezogene Terzkreis TR für die diatonischen Tonsets (a) C-Dur und (b) A-Dur.

4.2.6 Die diatonischen Tonleiterkreise D und DR

Analog zu den Kreisen F und FR bzw. T und TR handelt es sich bei Kreisen D und DR um den tonartübergreifenden bzw. tonartbezogenen diatonischen Tonleiterkreis. Sie repräsentieren die dritte durch die Symmetriemodell-Generatorformel erzeugte Ebene. In der gleichen Weise, wie der T aus dem F entstand, entsteht der D aus dem T . Abbildung 4.8 veranschaulicht, wie durch Addition einer Konstanten ± 6 zu jedem Element der Terzreihe die tonartübergreifende diatonische Tonleiterreihe entsteht. Wie beim F und T kann ein Bereich von 84 Halbtönen (12 Quinten) zu einem Kreis geschlossen werden, welcher als *tonartübergreifender diatonischer Tonleiterkreis* bezeichnet wird. Analog zum F und T repräsentiert ein um eine Symmetrieachse zentrierter Bereich von 12 Halbtönen

ein diatonisches Tonset und kann zum *tonartbezogenen diatonischen Tonleiterkreis* geschlossen werden. Die Kreise D und DR modellieren die Gruppierung von Tonigkeiten zu diatonischen Tonleitern. Der D enthält unter Annahme enharmonischer Equivalenz alle diatonischen Tonleitern in einer dem Quintenzirkel entsprechenden Anordnung. Auf eine Angabe der mathematischen Beschreibung wird an dieser Stelle verzichtet. Diese ergibt sich unter Berücksichtigung der Halbtonauflösungen $g_D = 84$ und $g_{DR} = 12$ analog zu denen im F und FR sowie T und TR . Der D und der DR für C -Dur sind in Abbildung 4.13 zu sehen.

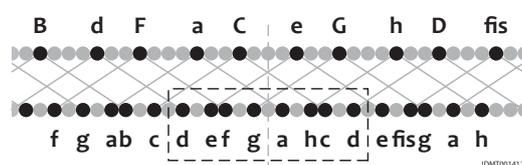


Abbildung 4.8 – Entstehung der diatonischen Tonleiterreihe aus der Terzreihe

4.2.7 Die chromatischen Kreise C und CR

Durch Addition einer Konstanten ± 3 zu jedem Element der diatonischen Tonleiterreihe ergibt sich die chromatische Reihe (Abbildung 4.9). Ein zu einem Kreis geschlossener Bereich von 84 Halbötönen (12 Quinten) ergibt den *tonartübergreifenden chromatischen Kreis C*, ein symmetrisch um die Symmetrieachse eines diatonischen Tonsets ($n_c = 7n_k$) verteilter Bereich von 12+1 Halbötönen kann zum *tonartbezogenen chromatischen Kreis CR* geschlossen werden. Durch die Transformation ± 3 kommt es in der chromatischen Reihe zu Doppelbelegungen, welche in Abbildung 4.9 durch kreisförmige Umrandungen gekennzeichnet sind. Die Kreise C und CR haben nach dem derzeitigen Stand im Vergleich zu den restlichen Teilräumen keine größere praktische Bedeutung und wurden hier nur der Vollständigkeit halber angegeben.

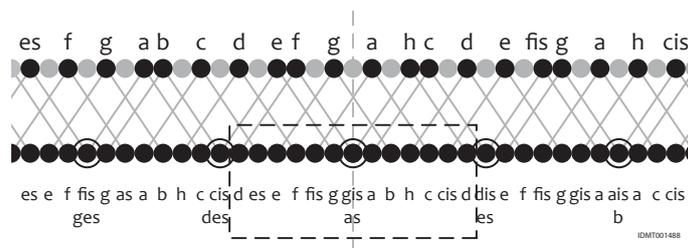


Abbildung 4.9 – Entstehung der chromatischen Reihe aus der diatonischen Tonleiterreihe

4.3 Erweiterungen des Symmetriemodells

4.3.1 Spiralförmige Darstellung des Symmetriemodells

Bei den Tonräumen *FR*, *TR* und *DR* handelt es sich um kreisförmig repräsentierte Ausschnitte aus der Quint-, Terz- und diatonischen Tonleiterreihe. Möchte man tonartbezogene und tonartübergreifende Aspekte gleichzeitig visualisieren, so bietet sich die spiralförmige Repräsentation dieser Modelle an. Hierbei werden die tonartbezogenen Kreise in der dritten Dimension nach unten und oben fortgeführt. Dies bedeutet, dass genau eine Spiralwindung ein diatonisches Tonset repräsentiert. In Abbildung 4.10 ist dies für den *T* bzw. *TR* in Form der *Terzspirale* dargestellt. Die schwarze Spiralwindung stellt hier den tonartbezogenen Terzkreis für C-Dur dar. Anfang und Ende des ursprünglichen Ausschnittes aus der Terzreihe bilden nun Anfang und Ende der Spiralwindung. Anwendungen der Terzspirale sind z.B. in Abschnitt 6.5.3 und Abschnitt 5.3.4 zu finden.

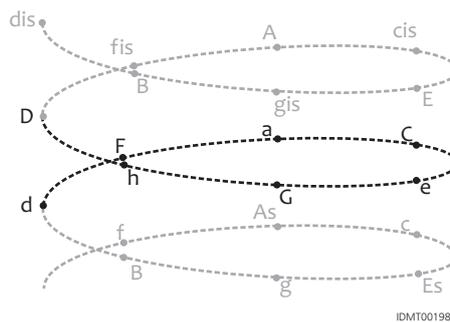


Abbildung 4.10 – Die Terzspirale

4.3.2 Repräsentation chromatischer Tonigkeiten im *FR*, *DR* und *TR*

In der tonalharmonischen Musik kommt es immer wieder vor, dass innerhalb einer diatonischen Tonart Klänge oder Tonigkeiten aus anderen Tonsets verwendet werden. Beispiele hierfür sind Alterationen (Abschnitt 5.3.6), Zwischendominanten (Abschnitt 5.4.3) oder chromatische Vorhalte. Die Besonderheit hierbei ist, dass sich das tonale Zentrum bzw. die wahrgenommene Tonart (Abschnitt 2.7.3) nicht ändern. D.h. auch die chromatischen nicht zu einem diatonischen Tonset gehörenden Tonigkeiten werden auf das aktuelle Zentrum bezogen. Wir müssen deshalb eine Möglichkeit finden, chromatische Tonigkeiten sinnvoll in ein gegebenes tonartbezogenes Modell einzuordnen. Eine Möglichkeit ist die Projektion der spiralförmigen Darstellung der Kreise *FR*, *TR* oder *DR* in die 2D-Ebene. In Abbildung 4.11a wird dies für den *TR* veranschaulicht: Die Terzreihe wird einfach zu

beiden Seiten der Symmetrieachse weiter geführt. Dadurch kommt es zu einer automatischen Zuordnung der chromatischen Tonigkeiten ..., *es*, *b*, *fis*, *cis* ... zu Hilfstonigkeiten im TR. Erhöhte Tonigkeiten werden im Uhrzeigersinn und erniedrigte Tonigkeiten gegen den Uhrzeigersinn von der Ausgangstonigkeit aus positioniert. In Abbildung 4.11b ist die Zuordnung der 14 chromatischen Tonigkeiten *ces* bis *eis* zum TR von C-Dur zu sehen. Es wird deutlich, dass der TR zwischen *enharmonischen Verwechslungen* einer Tonigkeit unterscheidet. So liegen z.B. die Tonigkeiten *dis* und *es* im TR gegenüber. Die hier beschriebene Erweiterung des Modells erlaubt eine bisher nicht mögliche Einordnung tonartfremder tonaler Elemente in ein tonartbezogenes Modell, so zum Beispiel die musiktheoretisch korrekte Einordnung des verminderten Septakkordes (Abbildung 5.18), des übermäßigen Sextakkordes (Abbildung 5.15) usw.

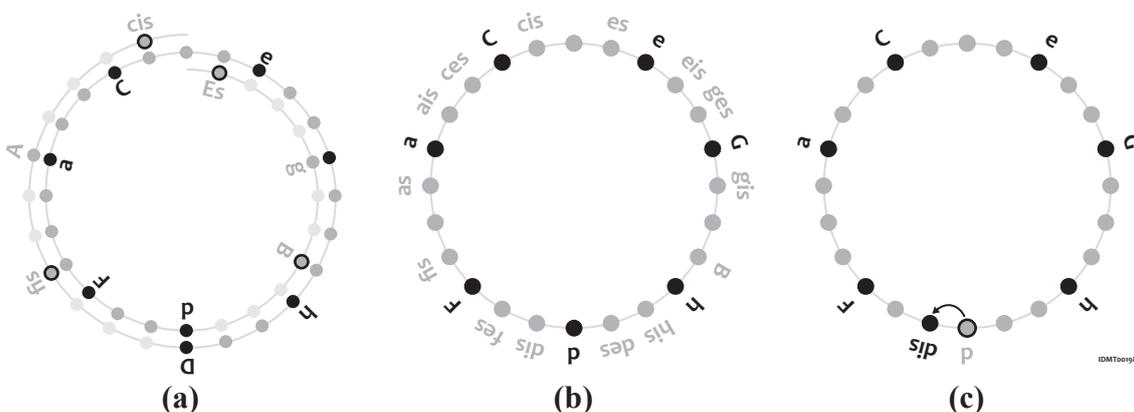


Abbildung 4.11 – Erweiterung des TRs zur Repräsentation nichtdiatonischer Tonigkeiten

4.3.3 Alterationen

Alterationen bezeichnen den Vorgang, dass einzelne Tonigkeiten eines Klanges chromatisch erhöht oder erniedrigt werden. Um Klänge mit alterierten Tonigkeiten in den tonartbezogenen Modellen analysieren zu können, führen wir die Möglichkeit ein, einzelne Tonigkeiten eines Tonraumes chromatisch zu erhöhen oder zu erniedrigen. Dies ist in Abbildung 4.11c am Beispiel des TRs und der zu *dis* hochalterierten Tonigkeit *d* veranschaulicht: Die Tonigkeit *dis* wird durch den Alterationsvorgang zu einer Realtonigkeit und wird deshalb schwarz ausgefüllt dargestellt. Die Ausgangstonigkeit *d* wird grau mit schwarzer Umrandung dargestellt, da diese zur Hilfstonigkeit geworden ist. Der Pfeil zwischen *d* und *dis* kennzeichnet darüber hinaus den Alterationsvorgang. Die Anwendung dieser Konventionen ist in Abschnitt 5.3.6 zu finden. Beim Alterieren von Tonigkeiten sind die Konventionen zur Einordnung nichtdiatonischer Tonigkeiten in die tonartbezogenen Räume des Symmetriemodells zu beachten (Abschnitt 4.3.2).

4.3.4 Repräsentation von Tonhöhe

Das Symmetriemodell, wie es bisher vorgestellt wurde, ist ein reines Tonigkeitsmodell (Abschnitt 2.3.3). Es fehlt bisher eine Möglichkeit, die unterschiedlichen Oktavlagen von Tonigkeiten zu repräsentieren. Eine Möglichkeit zur Repräsentation von Tonhöhe ist Wilhelm Moritz Drobischs Modell ([38], Abschnitt 2.3.3), welches die Tonigkeit in der XY-Ebene und die Tonhöhe auf der Z-Achse eines dreidimensionalen Koordinatensystems abbildet (Abbildung 2.2). Der Nachteil dieser Lösung ist jedoch das Hinzukommen einer dritten Dimension, was z.B. die Komplexität eines auf dem entsprechenden Tonraum basierenden Musikinstrumentes stark erhöhen würde. Analysiert man Drobischs Modell in einem Polarkoordinatensystem, so stellt man fest, dass in der XY-Ebene nur die Winkeldimension Träger von Information ist. Es ist deshalb möglich, die verbleibende radiale Dimension zur Repräsentation der Tonhöhe zu verwenden. Ergebnis wäre die Reduktion des dreidimensionalen Drobisch-Raumes auf einen *zweidimensionalen* Tonraum, welcher als *Tonhöhen-Tonigkeits-Raum* bezeichnet wird. Der Tonhöhen-Tonigkeits-Raum ist vor allem für die tonraumbasierte Synthese von Musik wichtig (Kapitel 6) und wird deshalb auch ausführlich in Abschnitt 6.3.1 erläutert.

4.4 Der Modulo- und der SYM-Operator

In diesem Abschnitt soll nun abschließend neben der Symmetriemodell-Generatorformel noch eine zweite Möglichkeit vorgestellt werden, wie die Teilräume des Symmetriemodells sowie eine Reihe weiterer Tonräume hergeleitet werden können. Abbildung 4.12 veranschaulicht die Grundidee. Zunächst sind hier verschiedene Spiralen zu sehen. Verfolgt man den Pfad entlang der Spiralwindung, so stellt man zunächst fest, dass hier in regelmäßigen Abständen Realtonigkeiten in Quintreihenfolge angeordnet sind. Der Unterschied zwischen den vier Abbildungen a, b, c und d besteht lediglich darin, dass sich unterschiedlich viele Hilfstonigkeiten (Halbtöne) auf einer Spiralwindung befinden. Durch diese Unterschiede kommt es zu einer bestimmten Zuordnung von Tonigkeiten über benachbarte Spiralwindungen hinweg.² In Abbildung 4.12 ist weiterhin ein diatonischer Ausschnitt aus der Quintreihe (sieben Quinten) hervorgehoben worden. Betrachtet man nur noch den Winkel der in diesem Ausschnitt enthaltenen Realtöne, so stellt man weiterhin fest, dass diese im Fall von Abbildung 4.12a, c, e und g genau den bereits eingeführten tonartbezogenen Räumen *F*, *FR*, *TR* und *DR* entsprechen. Nimmt man

² In Abbildung 4.12a sind z.B. die Tonigkeiten *his* und *c*, in Abbildung 4.12e die Tonigkeiten *c*, *e* und *g* oder in Abbildung 4.12g die Tonigkeiten *d*, *e* und *f* zusammengeordnet.

darüber hinaus nichtdiatonische Bereiche (grau dargestellt) hinzu, so erhält man zusätzlich die bereits in Abschnitt 4.3.2 besprochene Einordnung chromatischer Tonigkeiten in die Teilräume des Symmetriemodells. Es ist also möglich, die Anordnung von Tönen in den Teilräumen des Symmetriemodells herzuleiten, indem man eine Quintreihe auf einen Kreis mit unterschiedlichen Radien „aufwickelt“. Mathematisch kann dieser Prozess durch den Modulo-bzw. Rest-Operator beschrieben werden. Nehmen wir an, die Realtonigkeiten d, F, a, C, e, g und h im tonartbezogenen Terzkreis TR würden die Indizes $n_{TR} = 0, 3, 7, 10, 14, 17$ und 21 besitzen. Wir möchten nun die Tonigkeit f mit dem Index $n_F = -21$ (Abbildung 4.3) in den TR einordnen. Wir müssten dazu lediglich $n_F \bmod 24$ rechnen und erhalten den Wert $n_{TR} = 3$, was in Abbildung 4.12d genau der Tonigkeit f entspricht. Die Tonigkeit fis ($n_F = 28$) würde z.B. an der Stelle $28 \bmod 24 = 4$ in den TR eingeordnet werden. Dieser Wert identifiziert genau das um einen Halbton neben der Tonigkeit f ($n_{TR} = 3$) liegende fis . Diese Beispiele könnten für den FR und den DR fortgeführt werden. Von allgemeiner Bedeutung ist jedoch, dass die Anwendung von Modulo-Operationen auf eine Quintreihe eine Möglichkeit ist, verschiedene musikalische Zusammenhänge auszudrücken.

- Die Operation $n_F \bmod 84$ transformiert die Quintreihe so, dass enharmonisch identischen Tonigkeiten derselbe Zahlenwert zugeordnet wird (Abbildung 4.12a).³
- Die Operation $n_F \bmod 49$ ordnet Tonigkeiten mit gleichem Stammton (z.B. des, d, dis, \dots) den gleichen Wert zu (Abbildung 4.12b). Der daraus resultierende Tonraum wird deshalb vom Autor als Stammtonraum bezeichnet.
- Die Operation $n_F \bmod 48$ generiert einen Tonraum, der in der kreisförmigen Darstellung dem tonartbezogenen Quintenzirkel entspricht (Abbildung 4.12c) und dem Stammtonraum ($n_F \bmod 49$) sehr ähnlich ist. Unterschied zwischen letzterem und dem F ist die eindeutige Identifikation chromatischer Tonigkeiten.⁴
- Die Operation $n_F \bmod 28$ transformiert die Quintreihe so, dass Tonigkeiten die einen Abstand von einer großen Terz besitzen, derselbe Wert zugeordnet wird (Abbildung 4.12d). Der Kreis ist z.B. Bestandteil des 6D-Spaces von Hartes, Slanders und Gassers 6D-Space (Abschnitt 3.5). Die 3D-Repräsentation des Raumes ergibt Elaine Chews Spiral Array (Abschnitt 3.6).

³ Tonigkeiten wie z.B. gis ($n_F = -42$) und as ($n_F = +42$) wird durch die Modulo-84-Operation derselbe Wert (42) zugeordnet.

⁴ Den Tonigkeiten as ($n_F = -42$), a ($n_F = 7$) und ais ($n_F = 56$) werden durch die Modulo-Operation die Werte 6,7 und 8 zugeordnet, was die chromatische Benachbarung der Tonigkeiten wiedergibt.

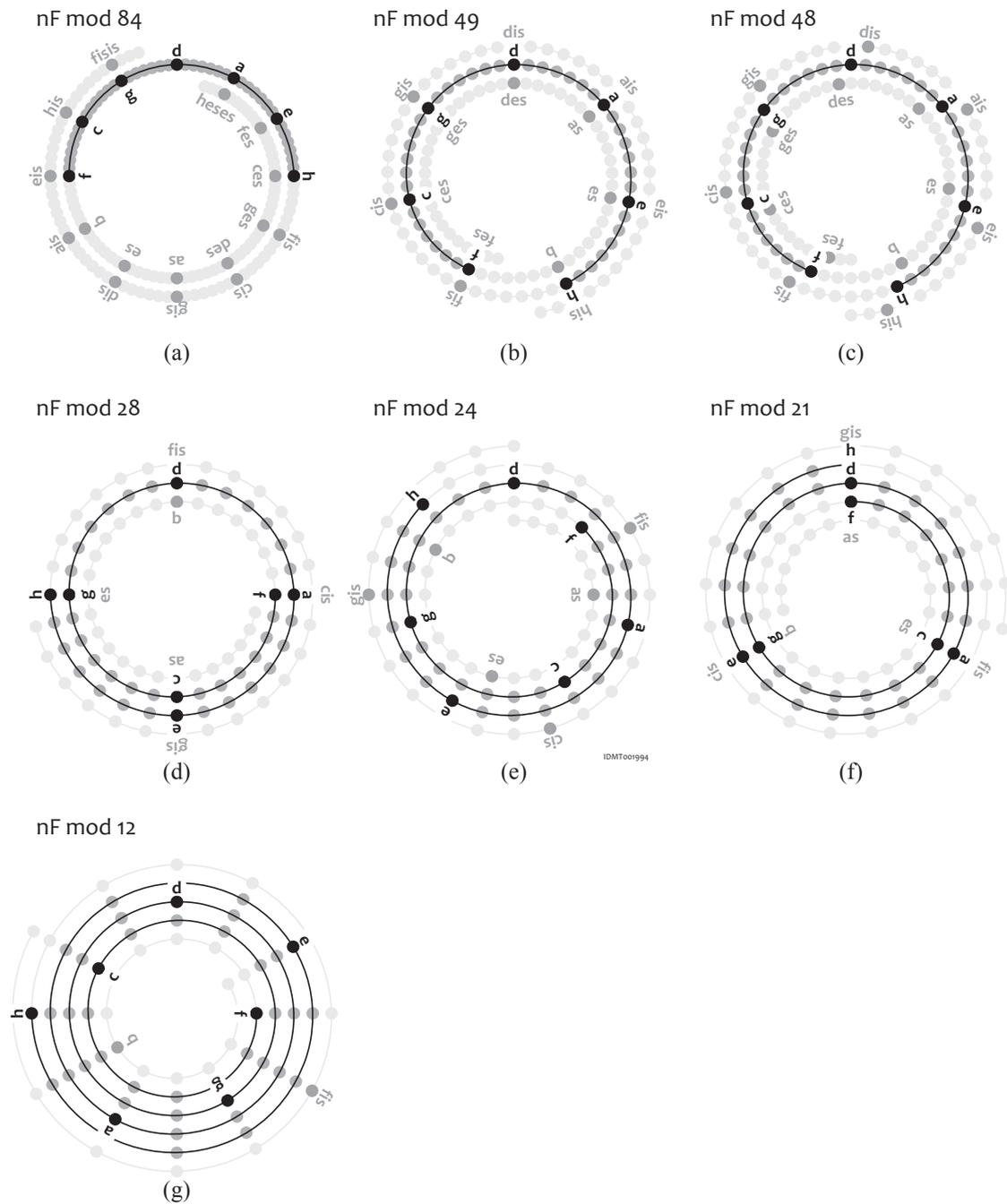


Abbildung 4.12 – Erzeugung verschiedener praxisrelevanter kreisförmiger Tonräume durch Anwendung des Modulo-Operators auf eine Quintreihe

- Die Operation $n_F \bmod 24$ ordnet Tonigkeiten zu Dur- und Moll- sowie verminderten Dreiklängen zusammen (Abbildung 4.12e).⁵ Die kreisförmige Repräsentation des resultierenden Raumes entspricht dem tonartbezogenen Terzkreis TR des Symmetriemodells (Abschnitt 4.2.5).
- Die Operation $n_F \bmod 21$ ordnet Tonigkeiten zusammen, welche einen Abstand von einer kleinen Terz besitzen (Abbildung 4.12f). Der resultierende Raum ist wie der Raum $n_F \bmod 28$ Bestandteil von Hartes, Sandler's und Gassers 6D-Space.
- Die Operation $n_F \bmod 12$ transformiert die Quintreihe so, dass Tonigkeiten entsprechend ihrer Reihenfolge in einer diatonischen Tonleiter repräsentiert werden (Abbildung 4.12g). Der resultierende Raum entspricht damit dem tonartbezogenen diatonischen Tonleiterkreis DR des Symmetriemodells (Abschnitt 4.2.6).

Wie anhand der vorherigen Auflistung zu sehen, ist es möglich, mit Hilfe des Modulo-Operators eine Quintreihe so „zusammenzufalten“, dass das Zahlenmaterial praxisrelevante Tonzusammenordnungen repräsentiert. Jedoch wird der Zahlenbereich auf Werte zwischen 0 und g abgebildet.⁶ Damit entsteht aber das Problem, dass die Symmetrieeigenschaften der abendländischen Tonalität und die damit zusammenhängenden Wahrnehmungseigenschaften sich nicht im numerischen Zahlenmaterial widerspiegeln. Es wird deshalb nun ein angepasster Modulo-Operator vorschlagen, welcher den Zahlenbereich auf Werte zwischen $-\frac{g}{2}$ und $+\frac{g}{2}$ abbildet und somit symmetrisch um den Wert 0 anordnet. Dieser Operator wird als SYM-Operator bezeichnet und ist wie folgt definiert:

$$m = \text{SYM}(n, g) = \left(n + \frac{g}{2}\right) \bmod g - \frac{g}{2} \quad (4.13)$$

Die Variable g ist die in Abschnitt 4.1.3 eingeführte Halbtonauflösung des zugrundeliegenden Kreissystems. Es gibt Fälle, in denen der Zahlenbereich nicht um den Wert 0, sondern um einen beliebigen Wert \tilde{n} gruppiert werden soll. Wir führen deshalb noch den dreiparametrischen SYM3-Operator ein:

$$m_{\tilde{n}} = \text{SYM}(n, g, \tilde{n}) = \text{SYM}(n - \tilde{n}, g) + \tilde{n} \quad (4.14)$$

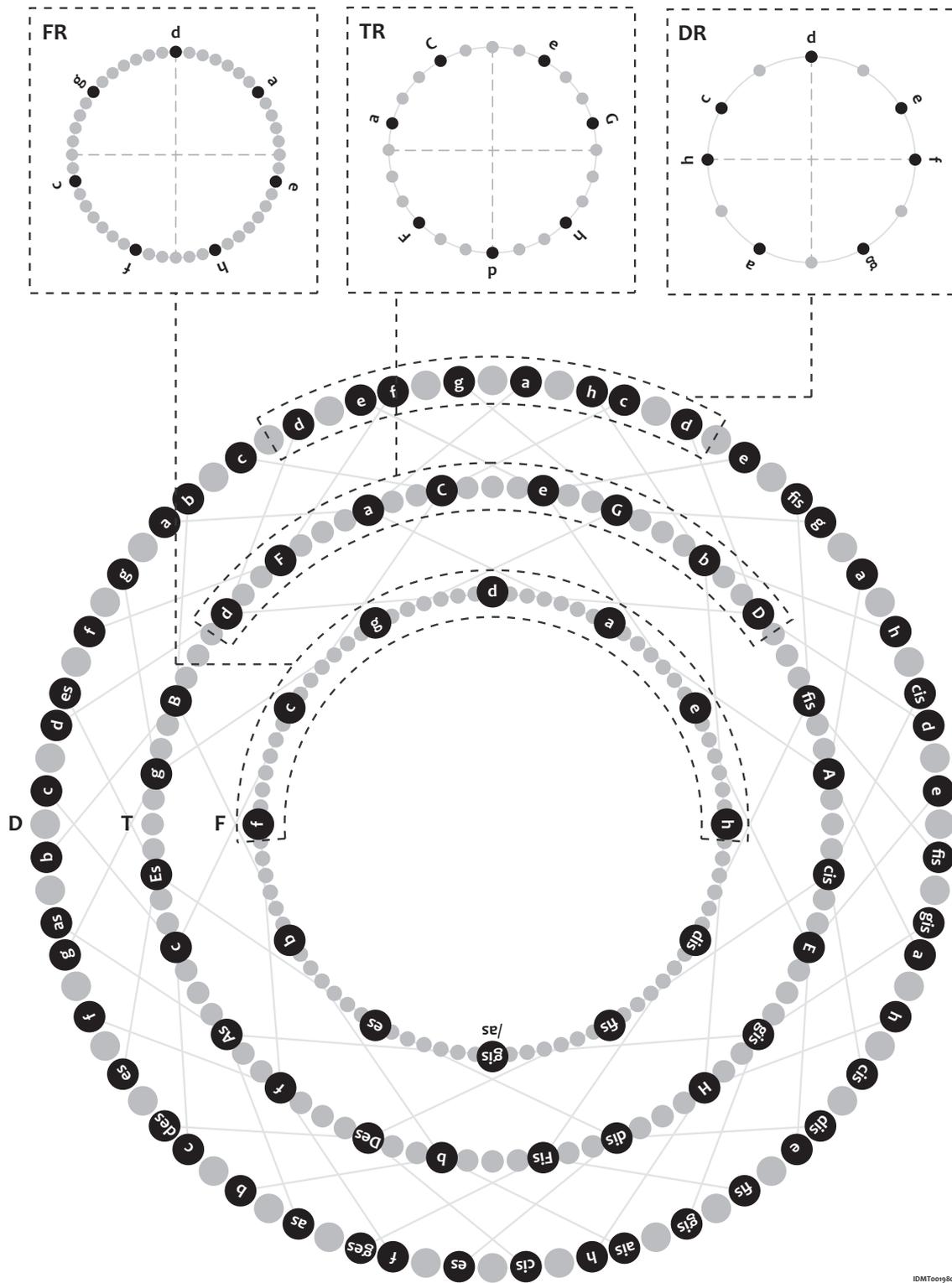
Anwendungsmöglichkeiten des SYM- und des SYM3-Operators sind:

- Die Einordnung einer chromatischen Tonigkeit n_C in die Quintreihe (Abschnitt B.3). Es gilt: $n_F = \text{SYM3}(49n_C, 84, 7n_k)$.

⁵ Z.B. werden die Töne c ($n_F = -14$), e ($n_F = 14$) und g ($n_F = -7$) durch die Modulo-Operationen auf die Werte 10, 14 und 17 abgebildet. Diese Zahlen spiegeln die Reihenfolge der Töne im Durdreiklang wider.

⁶ g bezeichnet die Halbtonauflösung des Kreises (siehe Abschnitt 4.1.3).

- Der Test, ob eine Tonigkeit n_F Bestandteil eines diatonischen Tonsets n_k ist (Abschnitt B.2). Es gilt: $-21 \leq n_F - 7n_k < +21$.
- Die Einordnung einer Tonigkeit n_F in einen der Unterräume des Symmetriemodells (Abschnitt B.3). Es gilt: $n_{FR} = \text{SYM}(n_F - 7n_k, 48)$, $n_{TR} = \text{SYM}(n_F - 7n_k - 12, 24)$, $n_{DR} = \text{SYM}(n_F - 7n_k, 12)$
- Der Test, ob eine Tonigkeit n_F in einem gegebenen diatonischen Tonset n_K Grundton eines Dur- oder eines Mollakkordes ist (Abschnitt B.4).
- Der Test, ob eine Tonigkeit n_F in einem gegebenen diatonischen Tonset n_K zum Subdominant-, Tonika- oder Dominantbereich gehört (Abschnitt B.4).
- Die Vorhersage des Auflösungsbestrebens einer Tonigkeit n_F in einer gegebenen Tonart (Abschnitt B.5).
- Die Berechnung der Tonigkeitsbezeichnung einer Tonigkeit n_F (Abschnitt B.6).
- Das Finden des enharmonisch äquivalenten Tonsets zwischen *Fis*-Dur und *Ges*-Dur zu einem beliebigen diatonischen Tonset mit n_k Vorzeichen (Abschnitt B.7).



IDMT00198a

Abbildung 4.13 – Das Symmetriemodell im Überblick

Kapitel 5

Darstellung musikalischer Elemente im Symmetriemodell

5.1 Einführung

Nachdem nun im vorherigen Kapitel die einzelnen Teilräume des Symmetriemodells eingeführt wurden, sollen nun häufig verwendete musikalische Elemente in das Symmetriemodell eingeordnet werden. Es wird gezeigt, dass viele dieser Elemente im Symmetriemodell Strukturen mit einfachen geometrischen Verhältnissen bilden. In der Tonart- und Akkorderkennung konnten durch den Einsatz des Symmetriemodells deutliche Verbesserungen erreicht werden. Das vorliegende Kapitel bietet den musiktheoretischen Hintergrund, um diese Dinge zu verstehen. Es ist wichtig voranzunehmen, dass alle der folgenden Darstellungen auf Basis idealer Tonigkeitskombinationen erfolgen. Damit ist in unserem Fall gemeint, dass die Tonigkeiten nicht durch harmonische Obertöne oder rauschartige Klangfarbenanteile gestört sind. Ideal meint in unserem Fall aber nicht, dass die Energie und damit die Hörbarkeit einer Tonigkeit nicht variieren könnten. Für eine Tonigkeitskombination wird angenommen, dass die Energie jeder Tonigkeit (Abschnitt 4.1.6) im Bereich zwischen 0,0 und 1,0 liegt. Die Analyse realer durch weitere Störtonigkeiten beeinflusster Tonigkeitskombinationen wird im Kapitel 10 behandelt.

5.1.1 Summenvektoroperationen im kreisförmigen Tonraum

Als Grundlage der nachfolgenden Ausführungen sollen nun noch einmal die bereits in Abschnitt 4.2 eingeführten grundlegenden Eigenschaften kreisförmiger Tonräume zusammengefasst werden:

- *Tonigkeiten und Tonigkeitsvektoren:* Jede in einem der Teilräume des Symmetriemodells repräsentierte Tonigkeit besitzt einen Winkel und einen Betrag. Der Betrag des Vektors beschreibt, wie stark die jeweilige Tonigkeit zu einem Zeitpunkt hörbar ist. Der Winkel beschreibt die Funktion der Tonigkeit hinsichtlich eines bestimmten Merkmals (siehe oben). Dieses ist von Teilraum zu Teilraum verschieden und geht aus der Reihenfolge der Tonigkeiten im Tonraum hervor.
- *Musikalisches Merkmal:* Die Teilräume des Symmetriemodells ordnen Tonigkeiten so an, dass Tonigkeiten, die hinsichtlich eines bestimmten Merkmals zusammengehören, sich in geometrischer Benachbarung befinden. Ein solches Merkmal kann z.B. ein Akkord, eine Tonart oder auch das Tongeschlecht sein. Eine Besonderheit der Teilräume F , FR , TR und DR besteht in der Tatsache, dass Tonigkeiten, die ein gemeinsames Merkmal teilen, etwa eine Halbebene belegen (Abbildung 5.1).
- *Summenvektor:* Durch die zuvor beschriebene Benachbarung zusammengehöriger Töne in den Teilräumen des Symmetriemodells ist es möglich, ein musikalisches Ereignis in Form eines Summenvektors zusammenzufassen: Die Vektoren aller Tonigkeiten werden addiert. Der Winkel des resultierenden Summenvektors beschreibt die Art des musikalischen Merkmals, welches durch das musikalische Ereignis repräsentiert wird. Die Länge des Summenvektors beschreibt, wie eindeutig das musikalische Merkmal ausgeprägt ist.¹
- *Merkmalunterstützende und merkmalabschwächende Tonigkeiten:* Hinsichtlich eines bestimmten musikalischen Merkmals kann die Menge aller Tonigkeiten in merkmalunterstützende und merkmalabschwächende Tonigkeiten unterschieden werden: Tonigkeiten, die ein Merkmal klar ausprägen, sind merkmalunterstützend. Tonigkeiten, die ein Merkmal verschleiern oder ein anderes Merkmal ausprägen, sind merkmalabschwächend. Zwei Tonigkeiten können dabei unterschiedlich stark merkmalunterstützend oder merkmalabschwächend wirken. Eine Tonart z.B. wird hauptsächlich durch die Stufen I , III und V (Grundton, die Terz und die Quinte) ausgeprägt, entfernter aber auch durch die die Stufen II , IV und VII . Wenn ein Tonraum gut gestaltet ist, dann sind zwei Tonigkeiten umso dichter beieinander positioniert, je stärker diese beiden Tonigkeiten ein bestimmtes Merkmal ausprägen.

Die vorher benannten Punkte werden in Abbildung 5.1 skizziert: Die Grafik veranschaulicht drei verschiedene musikalische Ereignisse, in denen jeweils drei Tonigkeiten hörbar

¹ Voraussetzung für diese Aussage ist eine Normierung des Audiosignals. Ansonsten würde sich natürlich auch die Intensität des Signals auf die Summenvektorenlänge auswirken.

sind. Es soll nun untersucht werden, ob die drei Ereignisse ein gegebenes Merkmal ausprägen. Dazu werden die Tonigkeiten in einen passenden Tonraum eingeordnet und entsprechend ihrer Hörbarkeit als Tonigkeitsvektoren repräsentiert (Abbildung 5.1, schwarze Pfeile). In Abbildung 5.1a ist nun der Fall zu sehen, dass die drei Tonigkeiten das zu analysierende Merkmal nur sehr schwach ausprägen: Die drei Tonigkeiten teilen sich in zwei geometrisch dicht beieinander liegende merkmalunterstützende Tonigkeiten auf und in eine dritte in einer anderen Halbebene merkmalabschwächende Tonigkeit. Der Summenvektor (grauer fetter Pfeil) ist deshalb sehr kurz. In Abbildung 5.1b ist der Summenvektor schon etwas länger, die beiden Tonigkeiten liegen in derselben Halbebene und etablieren dasselbe Merkmal. In Abbildung 5.1c kommen zu den zwei Tonigkeiten aus Abbildung 5.1b noch zwei weitere merkmalunterstützende Tonigkeiten hinzu, woraus eine weitere Verlängerung des Summenvektors resultiert.

Die nachfolgenden Betrachtungen unterliegen der Annahme, dass eine Tonigkeit entweder hörbar oder nicht hörbar ist. Die Länge des Vektors einer hörbaren Tonigkeit wird deshalb mit 1.0 und einer nicht hörbaren Tonigkeit mit 0.0 angenommen.

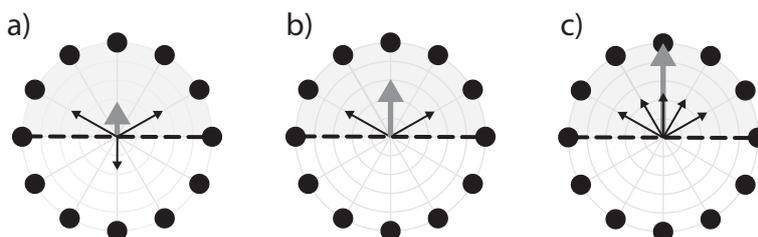


Abbildung 5.1 – Der Summenvektor als wichtiges Grundelement der Analyse musikalischer Elemente im Symmetriemodell

Auf Basis dieser Grundlagen werden nun sehr häufig vorkommende musikalische Elemente in das Symmetriemodell eingeordnet. Damit wird aufgezeigt, welche Merkmale durch die einzelnen Teilräume ausgeprägt bzw. unterdrückt werden.

5.2 Grundelemente

5.2.1 Einzelne Tonigkeiten

Abbildung 5.2 zeigt die Einordnung der Tonigkeit d in den tonartübergreifenden Terzkreis T und den tonartübergreifenden Quintenzirkel F . Im T ist die Tonigkeit einmal als Grundton des Akkordes d -Moll und einmal als Grundton des Akkordes D -Dur zu sehen (schwarze Pfeile). Möchte man z.B. die Tonigkeit d in den tonartübergreifenden Terzkreis T einordnen, so muss also zunächst bestimmt werden, in welchem der beiden Zusam-

wenn man den Kontext nicht berücksichtigt und die Tonigkeit mehrfach einordnet. Dies ist auch der Grund, warum die Kreise T , D und C in den nachfolgenden Ausführungen nicht weiter betrachtet werden. Die funktionstheoretische Einordnung von Tonigkeiten erfolgt in Abschnitt 5.3.1. Die tonale Stabilität von Tonigkeiten in einer gegebenen Tonart wird in Abschnitt 5.3.5 behandelt.

5.2.2 Intervalle

Da im Symmetriemodell keine Tonhöhen, sondern nur Tonigkeiten betrachtet werden, ist es nicht möglich, eine Unterscheidung zwischen Komplementärintervallen³ oder Intervallen größer als eine Oktave vorzunehmen. Zum Einbeziehen der Tonhöhe könnten die zweidimensionalen Teilräume des Symmetriemodells jeweils entsprechend Roger Shepards Spiral-Helix [153] oder des in Abschnitt 4.3.4 vorgeschlagenen Tonigkeits-Tonhöhenraumes um eine Tonhöhendimension erweitert werden. Die Summenvektoren von Intervallen in kreisförmigen Tonräumen und damit auch in den Teilräumen des Symmetriemodells haben folgende Eigenschaften:

- Bei Intervallen bewegt sich der Winkel des Summenvektors je nach Energieverhältnissen der beiden Tonigkeiten zwischen den Winkeln der Einzeltonigkeiten. Dies bedeutet: Je größer der Abstand zweier Tonigkeitsvektoren im Tonraum ist, desto größer ist auch die Überlappung unterschiedlicher Tonkombinationen und desto schlechter kann das betroffene Intervall von anderen unterschieden werden. In Abbildung 5.3 ist der potentielle Winkelbereich des Quintintervalls $c-g$ in Form eines grauen Kreissegmentes visualisiert. Das im jeweiligen Tonraum nächstliegende Quintintervall ist durch ein gestricheltes Kreissegment dargestellt. Die Überlappung der beiden Segmente variiert von Tonraum zu Tonraum stark.
- Vorausgesetzt die max. Energie von Tonigkeiten beträgt 1,0, dann bewegt sich die Länge des Summenvektors von Intervallen zwischen 2,0 (die beiden Tonigkeiten haben max. Energie und befinden sich an derselben Stelle im Tonraum⁴) und 0,0

³ „Das Komplementärprinzip ordnet die Intervalle gemäß ihrer gegenseitigen Ergänzung zur Oktave (lat.-fr. komplementär = sich gegenseitig ergänzend). Dabei ergänzen einander: rein und rein, klein und groß, übermäßig und vermindert, Prime und Oktave, Sekunde und Septime, Terz und Sexte, Quarte und Quinte, Sexte und Terz, Septime und Sekunde, Oktave und Prime“ [84, S. 7].

⁴ Dies ist z.B. bei der Prime der Fall. Hier überdecken sich die Tonigkeitsvektoren vollständig (siehe Abbildung A.1).

(die beiden Tonigkeiten haben gleiche Energie und liegen im Tonraum direkt gegenüber⁵).

- Der Summenvektor im kreisförmigen Tonraum hebt diejenigen Intervalle hervor, deren Tonigkeitsvektoren sich in einem Bereich kleiner 120° befinden. Im F sind dies die Intervalle Quinte, große Sekunde sowie kleine Terz, im FR die Quinte und die große Sekunde, im TR die Quinte sowie die kleine und die große Terz, im DR die kleine und große Sekunde sowie die kleine Terz. Die anderen Intervalle führen jeweils zu sehr kurzen Summenvektoren. Eine Gegenüberstellung aller Intervalle kann in Abbildung A.1 gefunden werden.
- Ein direkter Zusammenhang zwischen Intervallkonsonanz (Abschnitt 2.4.2) und Summenvektorlänge in den Kreisen F , FR , TR und DR ist nicht gegeben. Lediglich die Länge des Summenvektors im TR kommt der in 2.4a dargestellten Konsonanz von einfachen Sinustönen nahe.

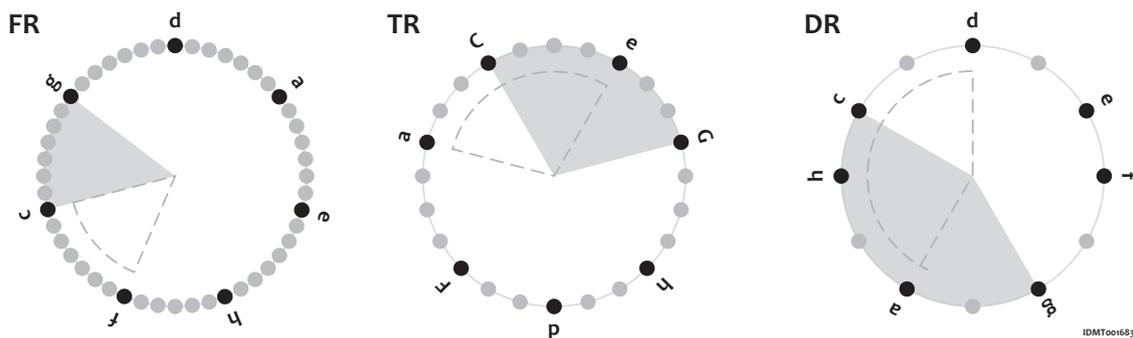


Abbildung 5.3 – Repräsentation benachbarter Quintintervalle in den Kreisen FR , TR und DR : Je näher die beiden Tonigkeitsvektoren eines Intervalls im Tonraum beieinander sind, desto eindeutiger kann das Intervall von anderen unterschieden werden.

5.2.3 Mehrklänge

Bei Mehrklängen, d.h. Kombinationen aus drei oder mehr Tonigkeiten, hängt der Winkelbereich des Summenvektors von der Verteilung der Tonigkeiten ab. Liegen die Tonigkeiten des Mehrklanges innerhalb eines Halbkreises, so bestimmen die äußeren Tonigkeiten den potentiellen Winkelbereich. Belegen die Tonigkeiten des Mehrklanges beide Halbebenen des Kreises, so kann der Winkel des Summenvektors jeden Winkel annehmen.

⁵ Dies ist z.B. beim Tritonus im Quintenzirkel der Fall. Hier liegen die Tonigkeiten h und f direkt gegenüber (Abbildung A.1, Teilraum F und C).

Ein Mehrklang führt dann zu einem langen Summenvektor, wenn dessen Tonigkeiten in Benachbarung liegen. Je stärker sie über den Winkelbereich des jeweiligen Kreises verteilt sind, desto kürzer wird der Summenvektor. Wie in Abschnitt 5.1.1 bereits beschrieben, werden die längsten Summenvektoren durch diejenigen Tonigkeiten gebildet, welche etwas weniger als einen Halbkreis belegen. Im *F* sind dies sechs benachbarte Tonigkeiten, in den Kreisen *FR*, *TR* und *DR* drei oder vier benachbarte Tonigkeiten.

Längen der Summenvektoren von Dreitonkombinationen

Dreitonkombinationen	Noten- beispiel	Summenvektorenlängen			
		F	FR	TR	DR/CR
Durdreiklang	c-e-g	1,93	0,80	2,21	0,52
Molldreiklang	a-c-e	1,93	0,80	2,21	0,52
Susakkord I	g-c-f	2,73	2,22	0,48	0,73
Verminderter Dreiklang	h-d-f	1,00	0,85	2,41	1,00
Übermäßiger Dreiklang	c-e-gis	0,00	0,52	1,73	1,00
Diatonische Dreitonfolge	d-e-f	1,51	0,61	0,87	2,40
Sonstige I	a-h-f	1,00	1,47	1,51	1,00
Sonstige II	a-c-d	2,39	1,36	1,47	1,50
Chromatische Folge	c-cis-d	0,73	1,22	0,52	1,73
Nichtdiatonische Melodiefolge	e-f-gis	0,52	1,32	0,26	1,93

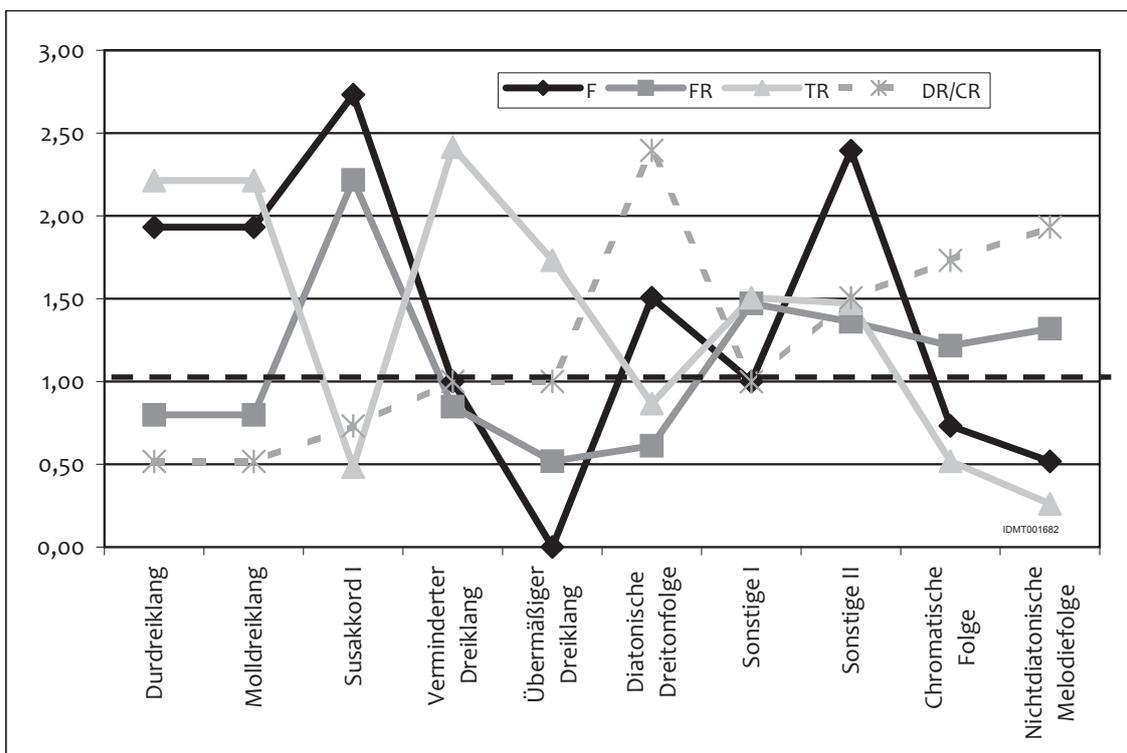


Abbildung 5.4 – Längen der Summenvektoren von Kombinationen aus drei Tonigkeiten in den Teilräumen des Symmetriemodells

- *Dur- und Molldreiklänge* werden im Symmetriemodell hauptsächlich durch den tonartbezogenen Terzkreis *TR* repräsentiert. In Abbildung 5.4 ist zu sehen, dass Dur- und Molldreiklänge im *TR* zu sehr langen Summenvektoren führen, also eine geringe Überlappung im Tonraum aufweisen und deshalb gut unterschieden werden können. Weiterhin ist zu sehen, dass mit Ausnahme des verminderten Akkordes alle anderen Dreitonkombinationen zu deutlich kürzeren Summenvektoren führen, was eine Unterscheidung von Dur- bzw. Molldreiklängen und anderen Dreitonkombinationen ermöglicht. Für die Unterscheidung von Dur- und Molldreiklängen spielt weiterhin der tonartbezogene Quintenzirkel *FR* eine wichtige Rolle: Obwohl es hier zu kurzen Summenvektoren kommt (Abbildung 5.4), trennt der *DR* Dur- und Molldreiklänge stark. Dies wird jedoch weiter unten noch genauer ausgeführt.
- *Vorhaltsakkorde*: In den Kreisen *F* und *FR* führen besonders Akkorde zu langen Summenvektoren, die sich durch Schichtung von Quinten beschreiben lassen. Dies sind z.B. der Sus4-Akkord⁶ oder der Durdreiklang mit hinzugefügter None. Die Spannungsgeladenheit dieser Akkorde führt dazu, dass sie in der Regel als *Vorhaltsakkorde* eingesetzt werden. Eine weitere Eigenschaft quintgeschichteter Akkorde ist die hohe Mehrdeutigkeit. So kann sich die Tonigkeitskombination *g-d-a* je nach Umkehrung allein in C-Dur in die Dreiklänge G-Dur, *a*-Moll oder *d*-Moll auflösen. Nimmt man noch leiterfremde Akkorde dazu, dann kommen hier noch die Akkorde *g*-Moll, *A*-Dur und *D*-Dur hinzu. Diese aus Akkordsicht hochgradige Mehrdeutigkeit hat zur Folge, dass der Susakkord im *TR* zu einem kurzen Summenvektor führt (Abbildung 5.4).
- Der *verminderte Dreiklang*⁷ hebt sich – was die Länge des Summenvektors betrifft – einzig und allein im tonartbezogenen Terzkreis *TR* hervor. In allen anderen Kreisen führt er zu kurzen Summenvektoren. Dies kann dahingehend interpretiert werden, dass der verminderte Akkord sowohl melodisch (*DR*) unbedeutend als auch hinsichtlich des Tongeschlechtes (*F* und *FR*) undeutlich ist. Funktionstheoretisch jedoch – was hauptsächlich durch den *TR* repräsentiert wird – spielt der verminderte Dreiklang eine wichtige Rolle, da er als Teil des Dominantseptakkordes (D^7), des verminderten Septakkordes (D^v), des Sixte ajoutée s^{65} (Moll) oder des subdominanten Sextakkordes s^6 (ebenfalls Moll) stark funktionsverstärkend wirkt [84].

⁶ Sus4-Akkorde sind Dur- oder Molldreiklänge, bei denen die Terz durch die Quarte ersetzt wurde.

⁷ Siehe Abschnitt 2.5.2

- *Der übermäßige Dreiklang*: Im tonartübergreifenden Quintenzirkel F heben sich die Tonigkeitsvektoren des übermäßigen Dreiklages komplett auf, was in einen Summenvektor der Länge 0 resultiert. Dies spiegelt den vagierenden Charakter [83, S. 46] des Dreiklages wider, was bedeutet, dass dieser Dreiklang in allen Tonarten zur Verwendung kommt und nur bei gegebener Tonart interpretiert werden kann. Im TR können nur zwei der drei Tonigkeiten des übermäßigen Dreiklages abgebildet werden, da die jeweils dritte Tonigkeit eine leiterfremde Tonigkeit ist⁸. Trotzdem kommt es hier zu langen Summenvektoren (Abbildung 5.4), was dahingehend interpretiert werden kann, dass der übermäßige Dreiklang in Form der Dominante mit hochalterierter Quinte eine starke funktionelle Wirkung hat [83, S. 46].
- *Diatonische Dreitonfolgen*, welche hauptsächlich für die Analyse melodischer Verläufe wichtig sind, führen im DR zu langen Summenvektoren (Abbildung 5.4). In den Kreisen TR und FR heben sich die Tonigkeitsvektoren weitgehend auf. Die Teilräume des Symmetriemodells besitzen damit die wichtige Eigenschaft, melodische und harmonische Komponenten eines Musikstückes zu trennen.
- *Sonstige diatonische und nichtdiatonische Tonigkeitsfolgen* werden, wie in Abbildung 5.4 zu sehen, in den Kreisen FR , TR und DR über Summenvektoren niederer und mittlerer Länge dargestellt. Sie ordnen sich damit zwischen Dur- und Molldreiklängen auf der einen Seite und melodischen Elementen auf der anderen Seite ein. Im tonartübergreifenden Quintenzirkel F variieren die Summenvektorenlängen sonstiger Tonigkeitsfolgen stark.

Sowohl für die semantische Audioanalyse als auch für die Erzeugung von Musik haben Dur- und Molldreiklänge große Bedeutung und sollen deshalb an dieser Stelle noch einmal separat in das Symmetriemodell eingeordnet werden. Abbildung 5.5 zeigt die Repräsentation des C-Dur-Dreiklages im Symmetriemodell. Für den Paralleldreiklang a -Moll ergibt sich eine an der vertikalen Symmetrieachse gespiegelte Repräsentation. Im TR belegen die Tonigkeiten des C-Dur-Dreiklages ($c-e-g$) einen zusammenhängenden Bereich und werden deshalb durch einen langen Summenvektor (2,21) repräsentiert. Im F sind zwar Grundton und Quinte ($c-g$) in Benachbarung, zwischen Terz (e) und Quinte (g) liegen jedoch zwei akkordfremde Tonigkeiten. Dies führt zwar zu langen Summenvektoren ($l = 1,93$), jedoch werden diese von quintgeschichteten Akkorden (z.B. $g-c-d$) deutlich übertroffen ($l = 2,73$, siehe auch Abbildung 5.4). Der F ist deshalb kein zuverlässiger Klassifikator für Dur- und Molldreiklänge. Im FR liegen zwar Grundton- und Quinte ($c-g$) in

⁸ Im Beispiel des Dreiklages $c-e-gis$ ist die Tonigkeit gis eine leiterfremde Tonigkeit.

Benachbarung, die Terz (e) liegt jedoch diesen beiden Tonigkeiten direkt gegenüber und verkürzt dadurch den Summenvektor stark ($l=0,8$). Im DR liegen die Tonigkeitsvektoren von Grundton und Quinte (c und g) relativ weit auseinander und führen dadurch zu kurzen Summenvektoren ($l=0,52$). Was die Trennung von Dur- und Molldreiklängen betrifft, so können nur durch eine Kombination der Modelle F , FR , TR und DR gute Ergebnisse erreicht werden. Der F ist geeignet tonal entfernte Akkorde zu trennen. Häufig zusammen auftretende quintverwandte Dreiklänge wie z.B. C-Dur und G-Dur oder auch parallelverwandte Dreiklänge wie z.B. C-Dur und a -Moll (siehe Abschnitt 5.3.4) werden dagegen schlecht getrennt. Die Summenvektoren liegen hier in einer Entfernung von 30° und damit dicht beieinander. Diese Schwäche wird jedoch durch die Kreise FR , TR und DR behoben. FR und DR separieren parallelverwandte Dreiklänge gut ($\Delta\alpha_{FR} = 166,9^\circ$, $\Delta\alpha_{DR} = 150^\circ$). Die Trennung von quintverwandten Dreiklängen wird dagegen besonders gut vom TR und vom DR vorgenommen ($\Delta\alpha_{TR} = 105^\circ$ und $\Delta\alpha_{DR} = 150^\circ$). Die Klassifikation von parallel- und quintverwandten Dreiklängen wird noch einmal in Abschnitt 5.3.1 im Zusammenhang mit Haupt- und Nebendreiklängen aufgegriffen, die Verwandtschaftsgrade von Dreiklängen sind Gegenstand von Abschnitt 5.3.4.

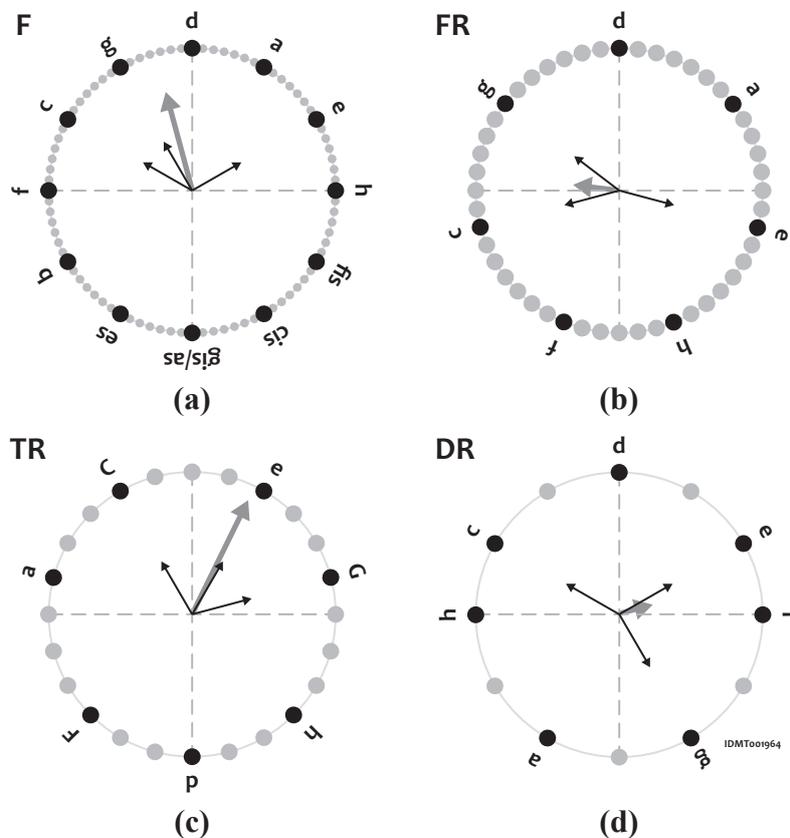


Abbildung 5.5 – Darstellung des C-Dur-Dreiklanges im Symmetriemodell.

5.2.4 Das diatonische Tonset

Das diatonische Tonset⁹ belegt in den tonartübergreifenden Modellen *F*, *T* und *D* einfach zusammenhängende Kreissegmente. Benachbarte diatonische Tonsets repräsentieren quintverwandte Tonarten. Abbildung 5.6 zeigt dies für die diatonischen Tonsets C-Dur (graue Kreissegmente) und F-Dur (gestrichelte Kreissegmente). Die Schnittmengenbereiche überlappender Kreissegmente repräsentieren gemeinsame Tonigkeiten unterschiedlicher Tonsets. Weiterhin wird ersichtlich, dass sich das zu einem diatonischen Tonset gehörende Kreissegment symmetrisch um die Symmetrieachse der gewählten Tonart anordnet. Der resultierende Summenvektor zeigt demnach genau auf die Stelle des leitereigenen Symmetrietones.

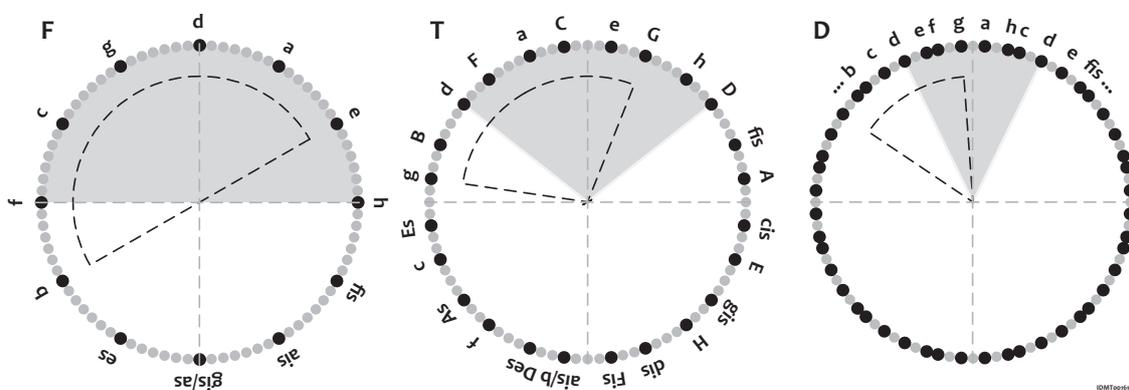


Abbildung 5.6 – Repräsentation des diatonischen Tonsets C-Dur (graues Segment) und F-Dur (gestricheltes Segment) in den Teilräumen *F*, *T* und *D* des Symmetriemodells (v. l. n. r)

Eine Besonderheit in der Darstellung diatonischer Tonsets ist der tonartübergreifende Quintenzirkel *F*. Wie in Abbildung 5.6a zu sehen ist, belegt das diatonische Tonset genau einen Halbkreis. Der Summenvektor im *F* ist deshalb um so länger, je mehr eine gegebene Menge von Tonigkeiten einem diatonischen Tonset entspricht. Dies ist in Abbildung 5.7 zu sehen, wo die Entwicklung der Länge des Summenvektors dargestellt ist, wenn zum Summenvektor nach und nach weitere Tonigkeitsvektoren addiert werden. Der Vektor wird länger, solange die Tonigkeiten Teil eines diatonischen Tonsets sind. Er verkürzt sich, wenn nichtdiatonische Tonigkeiten hinzukommen. Eine Ausnahme bilden die Tonigkeiten des Tritonus (*h-f*), welche gegenüber liegen und sich auslöschen.

⁹ Siehe Abschnitt 2.6.1

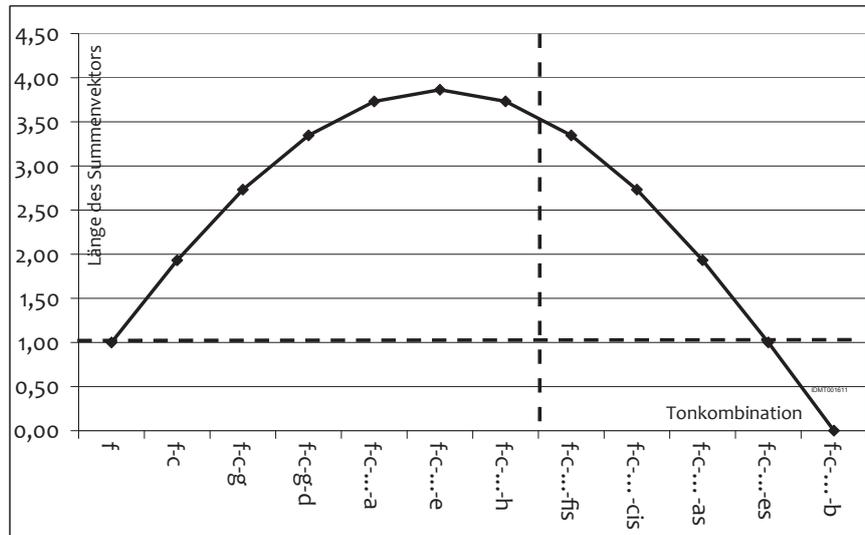


Abbildung 5.7 – Entwicklung der Länge des Summenvektors im F bei hinzukommenden Tonigkeiten. Die gestrichelte vertikale Linie trennt diatonische und nichtdiatonische Tonigkeiten.

5.3 Funktionstheoretische Eigenschaften

5.3.1 Haupt- und Nebendreiklänge, Haupt- und Nebenfunktionen

Die Funktionstheorie ordnet Akkorden sogenannte Funktionen zu. Die Funktion eines Akkordes hängt vom musikalischen Kontext ab und kann sich im Verlauf eines Musikstückes ändern. Mit der Funktion eines Akkordes sind die musikpsychologische Wirkung des Akkordes sowie die Übergangswahrscheinlichkeiten zu anderen Akkorden verknüpft. Zwei unterschiedliche Akkorde, welche die gleiche Funktion besitzen, können miteinander ausgetauscht werden, ohne dass die zugrundeliegende Akkordfolge ihre Schlüssigkeit verliert. Hier sind jedoch geltende Stimmführungsregeln zu beachten.

Bei Akkorden unterscheidet man Haupt- und Nebendreiklänge sowie Haupt- und Nebenfunktionen: Verwendet ein Musikstück vorrangig Durakkorde, dann werden die leitereigenen Durakkorde als *Hauptdreiklänge* und die leitereigenen Mollakkorde als *Nebendreiklänge* bezeichnet. Für Musikstücke in Moll ergibt sich eine umgekehrte Zuordnung von Haupt- und Nebendreiklängen. Die Funktion von Hauptdreiklängen wird als *Hauptfunktion*, die von Nebendreiklängen als *Nebenfunktion* bezeichnet. Die Hauptfunktionen heißen Tonika, Subdominante und Dominante und werden mit T , S und D (Dur) bzw. t , s und d (Moll) abgekürzt. Die *Tonika* ist der Dreiklang, „der im Zentrum aller harmonischen Spannungsfelder steht“ [84, S. 13]. Im *Symmetriemodell* wird diese Tatsache im tonartbezogenen Terzkreis TR veranschaulicht (Abbildung 5.8b): Hier befindet sich der

Summenvektor der Tonika *nah an der Symmetrieachse* und damit am geometrischen Zentrum des jeweiligen diatonischen Tonsets. Die *Subdominante* ist der Dreiklang, „der die untere Region der Tonika vertritt“ [84, S. 13]. Im *TR* wird diese Tatsache dadurch hervorgehoben, dass der *Summenvektor der Subdominante im negativen Winkelbereich*¹⁰ („unten“) zu finden ist. Die *Dominante* ist laut [84] der Dreiklang, „der die obere Region der Tonika harmonisch repräsentiert und dessen Kräfte von oben auf die Tonika einwirken“. Der *Summenvektor* der Dominante findet sich im Symmetriemodell *im positiven Winkelbereich* („oben“) wieder.

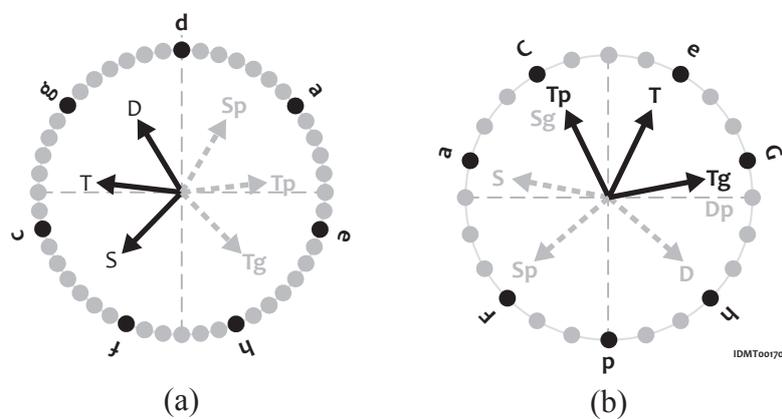


Abbildung 5.8– Haupt- und Nebendreiklänge werden (a) im tonartbezogenen Quintenzirkel *FR* getrennt und (b) im tonartbezogenen Terzkreis *TR* gruppiert (Beispiel für C-Dur). Aus Gründen der Übersichtlichkeit wurden nicht alle Funktionssymbole eingezeichnet.

Neben den Hauptdreiklängen und -funktionen gibt es noch die *Nebendreiklänge und -funktionen*. Ein Nebendreiklang kann einen Hauptdreiklang als *Klangvertreter* ersetzen und somit dessen Funktion übernehmen. Jeder Hauptdreiklang besitzt zwei leitereigene Klangvertreter, den sogenannten *Paralleldreiklang* und den *Gegendreiklang* [84]. So besitzt die Durtonika *T* (z.B. C-Dur) die Tonikaparallele *Tp* (*a*-Moll) und den Tonikagegenklang *Tg* (*e*-Moll). Im Symmetriemodell wird diese funktionelle Zusammengehörigkeit im tonartbezogenen Terzkreis *TR* veranschaulicht: Hier werden die Summenvektoren eines Hauptdreiklangs *T* und seiner Nebendreiklänge *Tp* und *Tg* in Benachbarung abgebildet (Abbildung 5.8b). Der tonartbezogene Quintenzirkel *FR* dagegen weist ein gegensätzliches Verhalten auf (Abbildung 5.8a): Hier befinden sich die Summenvektoren der Hauptdreiklänge zur einen Seite der Symmetrieachse und die Summenvektoren der Nebendreiklänge zur anderen. Der Summenvektor im *FR* sollte damit ein guter Klassifikator für Haupt- und Nebenfunktionen sein.

¹⁰ An dieser Stelle muss das verwendete Koordinatensystem (Abschnitt 4.1.1) beachtet werden.

5.3.2 Tonika-, Dominant- und Subdominantbereich

Ebenso wie ein Akkord eine bestimmte Funktion besitzt, je nachdem in welcher Tonart er vorkommt, so ist es auch möglich, einzelnen Tonigkeiten eine Funktion zuzuordnen, sofern bekannt ist, zu welchem Akkord und welcher Tonart sie gehören. Abbildung 5.9 teilt den tonartbezogenen Terzkreis für die Tonarten C-Dur und *a*-Moll in die drei Bereiche „Tonikabereich“, „Subdominantbereich“ und „Dominantbereich“ ein (Abbildung 5.9, *durchgezogene Pfeile*). Zum Tonikabereich werden dabei alle Tonigkeiten gezählt, die Bestandteil der Tonika sind. D.h. für die Tonart C-Dur zählen die Tonigkeiten *c*, *e* und *g* zum Tonikabereich, für *a*-Moll sind dies die Tonigkeiten *a*, *c* und *e*. Für den Subdominant- und Tonikabereich gilt der analoge Sachverhalt. Die in Abbildung 5.9 mit Hilfe durchgezogener Pfeile umrissenen Bereiche markieren die Zuordnung, wenn man von *reinen Dur- und Molldreiklängen* ausgeht. Betrachtet man Dur- und parallele Molltonart gemeinsam, so befinden sich die Tonikabereiche symmetrisch um die Symmetrieachse zentriert. Die Subdominantbereiche sowohl für Dur- als auch für Moll befinden sich links von der Symmetrieachse und die Dominantbereiche dazu gespiegelt rechts von der Symmetrieachse. Die *Überlappung der Bereiche* veranschaulicht die Tatsache, dass jede Tonigkeit mehrere funktionelle Bedeutungen haben kann, was sich auch in einer entsprechenden Mehrfachbedeutung von *Summenvektoren* widerspiegelt. Dem gemäß müssen statistische Modelle herangezogen werden, um die Summenvektoren in der richtigen Art- und Weise zu interpretieren.¹¹

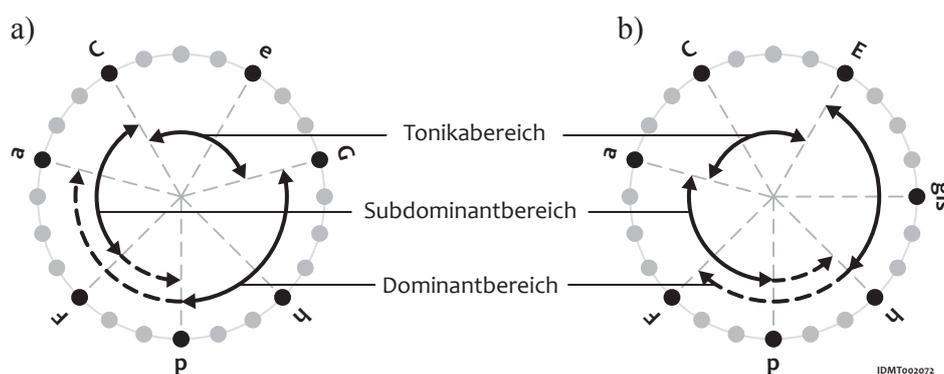


Abbildung 5.9 – Tonika-, Subdominant und Dominantbereich im tonartbezogenen Terzkreis: (a) für C-Dur und (b) für *a*-Moll

¹¹ Mehr dazu kann in Kapitel 10 gefunden werden.

5.3.3 Charakteristische Zusatzdissonanzen

Charakteristische Zusatzdissonanzen [84, S. 75] sind Tonigkeiten, die einem gegebenen Akkord hinzugefügt werden können, um dessen funktionelle Wirkung zu stärken bzw. eindeutig zu charakterisieren. Ein besonderes Merkmal hierbei ist, dass die Zusatzdissonanz aus dem Subdominantbereich und der betroffene Akkord aus dem Dominantbereich oder umgekehrt stammen. Die Eigenschaften charakteristischer Zusatzdissonanzen werden im *Symmetriemodell* insbesondere durch den tonartbezogenen Terzkreis modelliert: Wie in Abbildung 5.9 durch *gestrichelte Pfeile* visualisiert, liegen im tonartbezogenen Terzkreis *TR* Akkord und zugehörige charakteristische Zusatzdissonanz *in direkter Benachbarung*. Das Potential eines Akkordes einer gegebenen Funktion, eine Tonigkeit aus einem subdominantischen Funktionsbereich in den eigenen dominantischen Funktionsbereich (oder umgekehrt) „hinüberzuziehen“, wird geometrisch klar hervorgehoben. Die „herübergezogene“ Tonigkeit verstärkt die funktionelle Wirkung des „ziehenden“ Akkordes. Gleichzeitig erweitern die Zusatzdissonanzen damit die entsprechenden in Abschnitt 5.3.2 definierten Funktionsbereiche (Abbildung 5.9, gestrichelte Erweiterungen). Nachfolgenden werden drei Beispiele für charakteristische Zusatzdissonanzen genannt und auf die entsprechenden Detailkapitel verwiesen.

- *Der Dominantseptakkord*: Durch Hinzufügen der eigentlich subdominantischen Charakterdissonanz „kleine Septime“ zur Durdominante¹² wird die dominantische Wirkung dieses Dreiklanges verstärkt und „geschärft“ [84, S. 75]. Eine ausführliche Betrachtung des Dominantseptakkordes im Kontext des Symmetriemodells ist in Abschnitt 5.4.2 zu finden.
- *Der Sixte ajoutée*: Durch Hinzufügen der dominantischen Charakterdissonanz „Sexte“ zur Moll- oder Durdominante¹³ entsteht ein charakteristischer Vierklang, der sogenannte Sixte ajoutée [84, S. 113]. In diesem Fall verstärkt die Sixte den subdominantischen Charakter des Akkordes. Eine ausführliche Betrachtung des Sixte ajoutée im Kontext des Symmetriemodells ist in Abschnitt 5.4.4 zu finden.¹⁴
- *Der Dominantnonenakkord*: Fügt man dem Dominantseptakkord auch noch die None

¹² Im Beispiel aus Abbildung 5.9 wird der Durdominanten *G-h-d* die kleine Septime „F“ hinzugefügt.

¹³ Im Beispiel aus Abbildung 5.9 wird der Mollsubdominanten „d-F-a“ die Sixte „h“ bzw. der Dursubdominanten „F-a-C“ die Sixte „d“ hinzugefügt.

¹⁴ Weitere Beispiele für charakteristische Zusatzdissonanzen sind sowohl die subdominantischen Sextakkorde [84, S. 113] als auch der Dominantnonenakkord [84, S. 91].

hinzu, so entsteht der Dominantnonenakkord.¹⁵ Ähnlich wie auch die subdominante kleine Septime auf die dominante Seite „gezogen“ wird, wird auch die subdominante None auf die Seite der Dominante gezogen. Sie führt ebenfalls zu einer Verstärkung des funktionellen Charakters des Akkordes.

5.3.4 Verwandtschaftsgrade von Akkorden

Um die in Musikstücken vorkommenden Akkordverbindungen charakterisieren zu können, wurden verschiedene Verwandtschaftsgrade von Akkorden definiert. Die Basis dieser Verwandtschaftsgrade ist zum einen physikalischer Natur: So werden zwei Akkorde, deren Töne in einfachen Frequenzverhältnissen stehen, oder zwei Akkorde, die gemeinsame Töne haben, als verwandt wahrgenommen. Wesentlichen Einfluss auf die wahrgenommene Verwandtschaft von Akkorden hat auch der tonale Kontext (Abschnitt 2.7). Dieser bedingt zum einen die wahrgenommene Stabilität eines Akkordes (Abschnitt 2.7.1) als auch die wahrgenommene Nähe zweier Akkorde (Abschnitt 2.7.2). Zum anderen spielt auch eine kognitive Komponente eine wichtige Rolle bei Verwandtschaftsgraden. Ein Beispiel hierfür sind der Dur- und sein gleichnamiger Mollakkord: Die Grundtöne dieser Akkorde stehen in einem sehr einfachen Frequenzverhältnis (Prime), die innere Intervallstruktur ist sehr ähnlich (Terz-Quinte), und trotzdem haben beide Akkorde einen sehr unterschiedlichen emotionalen Affekt. Im Folgenden sollen verschiedene tonartübergreifende Verwandtschaftsgrade von Akkorden im Kontext des Symmetriemodells behandelt werden. Insbesondere soll dabei gezeigt werden, dass die in der Theorie definierten Verwandtschaftsgrade im Symmetriemodell in sehr einfachen geometrischen Verhältnissen repräsentiert werden.

Bei den Verwandtschaftsgraden von Akkorden unterscheidet man die Quint- und die Terzverwandtschaften, wobei sich die Begriffe „Quint“ und „Terz“ auf die physikalischen Intervallabstände der Akkordgrundtöne beziehen. Die *Quintverwandtschaften* sind das „statische Grundmuster der Musik zwischen 1600 und 1900“ [83, S. 29]“ und finden sich in Form der Hauptdreiklänge Tonika, Dominante und Subdominante wieder (Abschnitt 5.3.1). Bei den *Terzverwandtschaften* gibt es die leitereigenen Terzverwandtschaften sowie die entfernten Terzverwandtschaften 1., 2. und 3. Grades. Die *leitereigenen Terzverwandtschaften* kommen in Form der in Abschnitt 5.3.1 behandelten Nebendreiklänge vor. Die entfernten *Terzverwandten 1., 2. und 3. Grades* werden besonders einfach in der *Terzspirale* des Symmetriemodells (Abschnitt 4.3.1) repräsentiert. Wie in Abbildung 5.10

¹⁵ Im Beispiel aus Abbildung 5.9 wird dem Dominantseptakkord *G-h-d-F* die Tonigkeit „a“ als None hinzugefügt.

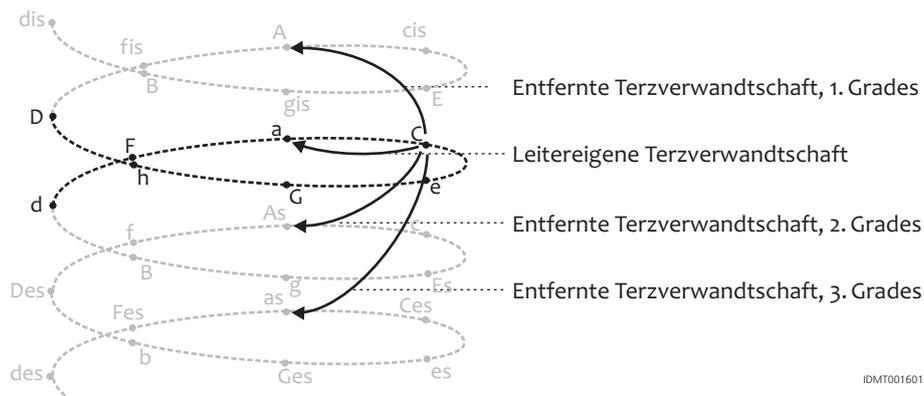


Abbildung 5.10 – Repräsentation leitereigener und entfernter Terzverwandtschaften in der Terzspirale am Beispiel der Parallelverwandtschaften des Dreiklages C-Dur

zu sehen, liegen die Terzverwandten in der Terzspirale *direkt übereinander*, d.h. sie besitzen dieselbe Z-Koordinate: So liegt direkt über der leitereigenen Terzverwandten die entfernte Terzverwandte 1. Grades. Direkt darunter liegt die entfernte Terzverwandte 2. und 3. Grades. Angewendet werden diese Zusammenhänge z.B. für tonraumbasierte Musikinstrumente (Kapitel 6). Weitere Aspekte der Verwandtschaft von Akkorden können gefunden werden in Abschnitt 2.1, Abschnitt 2.7.2 und Abschnitt 3.4.

5.3.5 Der Leitton, Strebetonigkeiten und Strebefähigkeit

Die *Strebefähigkeit* ist das Potential einer Tonigkeit [83, S. 45], sich in eine andere melodisch benachbarte Tonigkeit weiterzuentwickeln. Die strebende Tonigkeit wird im Folgenden als *Strebetonigkeit* und die folgende Tonigkeit als *Auflösungstonigkeit* bezeichnet. Je größer die Strebefähigkeit einer Tonigkeit ist, desto höher ist die Wahrscheinlichkeit, dass der Tonigkeit die zugehörige Auflösungstonigkeit folgt. Das Erklingen einer Strebetonigkeit baut demnach eine große Spannung auf. Die Eigenschaften von Strebetonigkeiten können gut im *TR* beschrieben werden. Hier liegen die potentiellen Strebetonigkeiten im Rücken des Tonikabereiches. In Abbildung 5.11 ist dies für das diatonische Tonset C-Dur dargestellt. In diesem Fall sind die Strebetonigkeiten die Tonigkeiten *h* und *F*. Die Auflösungstonigkeiten liegen im Bereich des geometrischen Zentrums und damit im Tonikabereich (Abschnitt 5.3.2). In Abbildung 5.11 sind dies die Tonigkeiten *C* und *e*. Die zu einer Strebetonigkeit gehörende Auflösungstonigkeit ist die im *TR* von der Strebetonigkeit am weitesten entfernte und damit annähernd im Kreis gegenüberliegende Tonigkeit. Im Abbildung 5.11 ist dies durch Pfeile zwischen den Tonigkeiten visualisiert: Dort strebt die Strebetonigkeit *h* in die Auflösungstonigkeit *C* und die Strebetonigkeit *F* in die Auflösungstonigkeit *e*. Der Symmetrietonigkeit *d* liegen die zwei Tonigkeiten *C* und *e*

Moll“ (Abschnitt 5.5.2) ein Strebeton in den Molltonikagrundton geschaffen. In Abbildung 5.12c wird die Tonigkeit *g* nach *gis* erhöht, was eine Strebetonigkeit nach *a* schafft.¹⁷ 4.) die *Erniedrigung der großen Durdominantnone*: Hier wird ein Strebeton in Richtung der Tonikaquinte geschaffen. In Abbildung 5.12d wird die Tonigkeit *a* nach *as* erniedrigt, was eine Strebetonigkeit nach *g* schafft.¹⁸

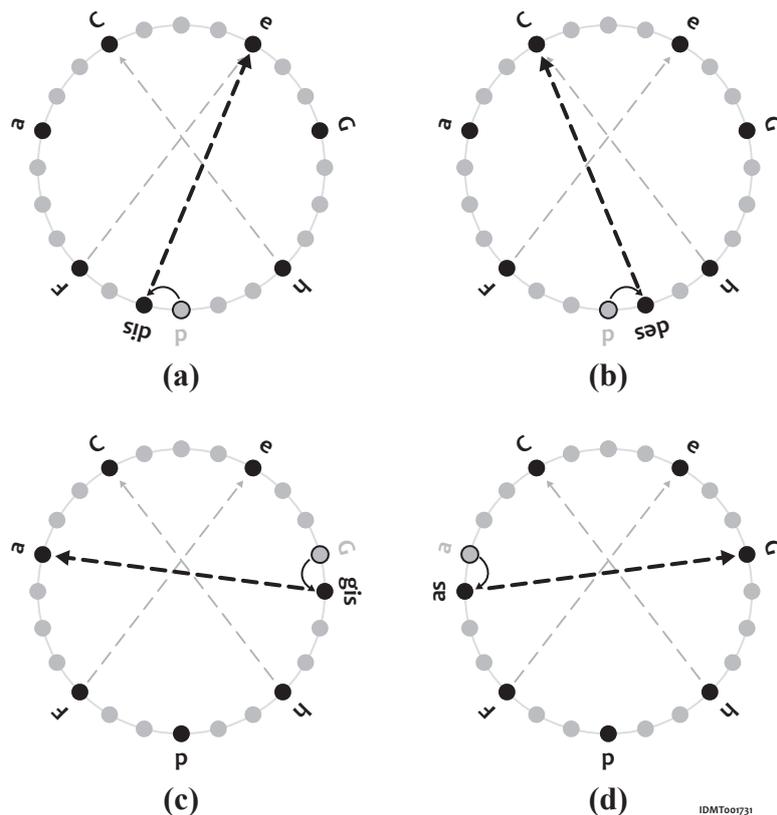


Abbildung 5.12 – Wichtige Alterationen ergänzen die bestehenden Strebetonigkeiten eines diatonischen Tonsets (hell gestrichelte Pfeile) um weitere Strebetonigkeiten (fette, schwarz gestrichelte Pfeile)

¹⁷ Siehe auch Abschnitt 5.5.2 (Harmonisch Moll) und Abschnitt 5.4.5 (Verminderter Septakkord).

¹⁸ Siehe auch Abschnitt 5.4.5 (Verminderter Septakkord).

5.4 Charakteristische Akkorde

5.4.1 Der verminderte Dreiklang

Der verminderte Dreiklang wird durch die im Rücken des TRs liegenden Tonigkeiten beschrieben¹⁹. Der Charakter des verminderten Dreiklages ist hochgradig instabil, was sich in seiner Mehrdeutigkeit und in seinen Deutungsmöglichkeiten bemerkbar macht. Derartig instabile Akkorde wurden von Arnold Schönberg als *vagierende Akkorde* bezeichnet, was noch im Zusammenhang mit dem verminderten Septakkord genauer erläutert wird (Abschnitt 5.4.5). Im Symmetriemodell kommen die Mehrdeutigkeiten bzw. Instabilitäten des verminderten Dreiklages und damit auch verwandter Dreiklänge wie dem Dominantseptakkord (Abschnitt 5.4.2) oder dem verminderten Septakkord (Abschnitt 5.4.5) klar zum Vorschein. Beispiele hierfür sind:

- Die funktionelle Mehrdeutigkeit des verminderten Dreiklages: Die Tonigkeiten des verminderten Dreiklages können sowohl subdominantisch als auch dominantisch gedeutet werden.²⁰ Wie im tonartbezogenen Terzkreis *TR* (Abbildung 5.13c) gesehen werden kann, überdeckt der Verminderte hier sowohl Teile des Dominant- als auch Teile des Subdominantbereiches (siehe auch Abschnitt 5.3.2). Der *Summenvektor* des Verminderten zeigt genau an die Stelle der Symmetrieachse, das ist die Grenze zwischen Subdominant- und Dominantbereich. Im tonartbezogenen Terzkreis *TR* wird die Zusammengehörigkeit des Subdominant- und des Tonikateils – obwohl funktionell so gegensätzlich – durch die geometrische Benachbarung gut dargestellt.
- Die Mehrdeutigkeit des verminderten Dreiklages hinsichtlich des Tongeschlechts (Abschnitt 5.5.1): Die Tonigkeiten können sowohl als Dur- als auch als Mollakkord interpretiert werden.²¹ Wie im tonartbezogenen Quintenzirkel *FR* (Abbildung 5.13b) gesehen werden kann, überdeckt der Verminderte hier den Moll- und den Durbereich (siehe auch Abschnitt 5.5.1). Der *Summenvektor* zeigt hier ebenfalls genau an die Stelle der Symmetrieachse, das ist die Grenze zwischen Dur- und Mollbereich.

¹⁹ Im Beispiel der Tonart C-Dur sind dies die Tonigkeiten *h*, *d* und *F*.

²⁰ Zum Beispiel können die Tonigkeiten *h-d-F* im Kontext der Tonart *a*-Moll als subdominantischer Sextakkord *d-F-h* (*s*⁶) [83, S. 121] und im Kontext der Tonart C-Dur auch als verkürzter Dominantseptakkord (*G*)-*h-d-f* gedeutet werden.

²¹ Zum Beispiel können die Tonigkeiten *h-d-F* im Kontext von *a*-Moll als Moll-Sextakkord (*d-F-h*) und im Kontext von C-Dur als Dur-Septakkord ohne Grundton (*G*)-*h-d-F* gedeutet werden.

- Die Offenheit des verminderten Dreiklages zu benachbarten Tonarten: Schaut man sich den Verminderten im tonartübergreifenden Terzkreis (Abschnitt 4.2.4) oder auch in der Terzspirale an (Abschnitt 4.3.1), so sieht man, dass er sowohl Anfang als auch Ende eines diatonischen Bereiches und damit einer Spiralwindung bildet. Er bietet damit eine Schnittstelle zu benachbarten Tonarten.
- Eindeutige Zuordnung des verminderten Dreiklages zu einem diatonischen Tonset: Im tonartübergreifenden Quintenzirkel F (Abbildung 5.13a) belegen die Tonigkeiten des verminderten Dreiklages genau eine Halbebene, welche exakt einem diatonischen Tonset zugeordnet ist (Abschnitt 5.2.4). Der *Summenvektor* in F zeigt im Falle des verminderten Dreiklages direkt auf die Symmetrietonigkeit dieses diatonischen Tonsets.
- Hohe Strebefähigkeit des verminderten Dreiklages in Richtung der Tonika: Entsprechend Abbildung 5.11a weisen zwei der drei Tonigkeiten des verminderten Dreiklages eine starke Strebewirkung in Richtung der Tonika auf (siehe auch Abschnitt 5.3.5). Der *Summenvektor* des Akkordes im TR hat die maximale Entfernung vom geometrischen Zentrum des leitereigenen diatonischen Tonsets, was in Anlehnung an Abschnitt 5.5.4 die Instabilität des Dreiklages repräsentiert.
- Die Mehrdeutigkeit des verminderten Dreiklages hinsichtlich der Strebefähigkeit (Abschnitt 5.3.5): Zwei der drei Tonigkeiten des verminderten Dreiklages²² sind gleichzeitig auch Strebetonigkeiten der im Quintenzirkel gegenüberliegenden Tonart. Hier führt der verminderte Dreiklang im TR auch zu einem relativ langen Summenvektor, welcher ebenfalls weit vom tonalen Zentrum entfernt ist und damit die hohe Strebefähigkeit des verminderten Dreiklages repräsentiert.

Während im TR des Symmetriemodells (Abbildung 5.13c) die *funktionelle* Mehrdeutigkeit des verminderten Dreiklages deutlich hervorgehoben wird, tritt im tonartübergreifenden Quintenzirkel F die *diatonische* Mehrdeutigkeit des verminderten Dreiklages in den Vordergrund: Wie in Abbildung 5.13a zu sehen, belegen die Tonigkeitsvektoren des verminderten Dreiklages einen kompletten Halbkreis, was auch die *Offenheit des Akkordes zu anderen Tonarten* verdeutlicht. Diese Offenheit führt dazu, dass der verminderte Dreiklang z.B. als Bestandteil des Dominantseptakkordes (Abschnitt 5.4.2) oder des verminderten Septakkordes (Abschnitt 5.4.5) in *verschiedensten Tonarten* zum Einsatz kommt und deshalb auch ein wichtiges *Modulationsmittel* darstellt (siehe auch Abschnitt 5.5.3).

²² Im Beispiel des verminderten Dreiklages $h-d-F$ sind dies die Tonigkeiten h und F , welche als H und eis Strebetonigkeiten der Tonart Fis -Dur und als Ces und f Strebetonigkeiten der Tonart Ges -Dur sind.

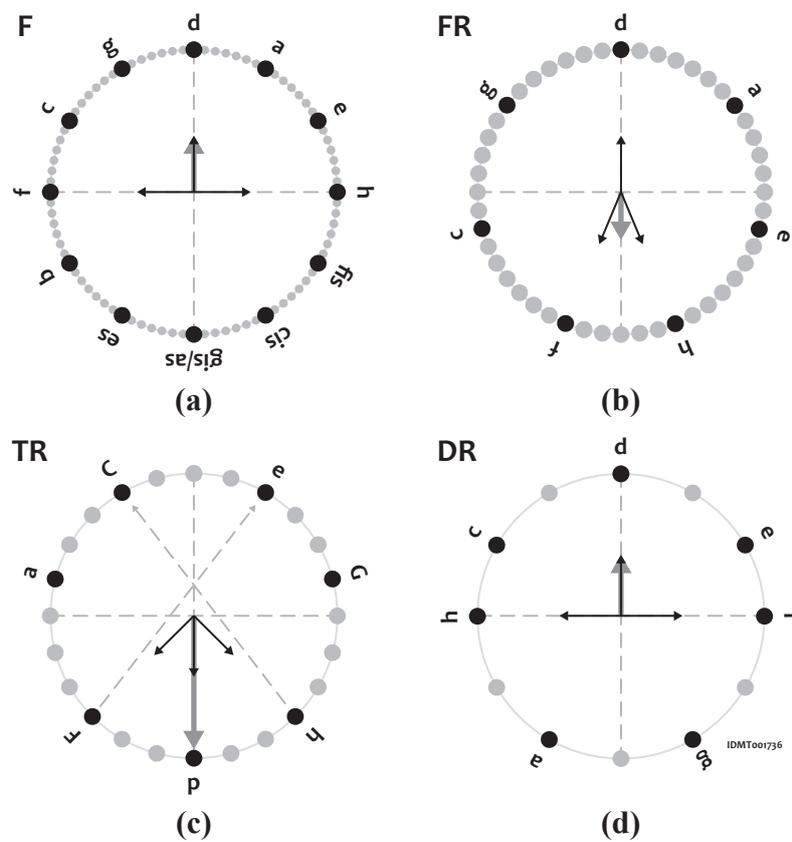


Abbildung 5.13 – Repräsentation des verminderten Dreiklages in den Teilräumen des Symmetriemodells: (a) *F*, (b) *FR*, (c) *TR* und (d) *DR*

5.4.2 Der Dominantseptakkord

Fügt man einem Durakkord die kleine Septime hinzu, so erhält man den Dominantseptakkord. Dieser Vierklang hat für die tonalharmonische Musik eine sehr große Bedeutung. Der Dominantseptakkord enthält einen verminderten Dreiklang (siehe Abbildung 5.14) und teilt deshalb mit diesem eine Reihe von Eigenschaften. Zu diesen Eigenschaften gehören die in Abschnitt 5.4.1 besprochene funktionelle Mehrdeutigkeit, die eindeutige Zuordnung des Akkordes zu einem diatonischen Tonset, die hohe Strebefähigkeit des Dominantseptakkordes in Richtung der Tonika und auch die Mehrdeutigkeit hinsichtlich der Strebefähigkeit²³. Diese Eigenschaften sollen an dieser Stelle nicht noch einmal diskutiert werden. Unterschiede des Dominantseptakkordes im Vergleich zum verminderten Dreiklang sind vor allem die Eindeutigkeit hinsichtlich der Funktion: Im Gegensatz zum mehrdeutigen verminderten Dreiklang ist die Funktion des Durakkordes mit hinzugefügter kleiner Septime ohne Kontext eindeutig dominantisch. Das wird auch durch den *Summenvektor* im *TR* widergespiegelt, welcher klar im Dominantbereich liegt. Der subdominante Teil des Dominantseptakkordes schwächt nicht dessen dominante Wirkung ab, auch wenn man es vermuten könnte. Im Gegenteil wird die dominante Wirkung noch verstärkt. Näheres dazu wurde in Abschnitt 5.3.3 im Zusammenhang mit Charakterdissonanzen ausgeführt.

Umdeutungsmöglichkeiten des Dominantseptakkordes

Der Dominantseptakkord enthält den verminderten Dreiklang und damit sehr viele instabile Tonigkeiten. Eine besondere Eigenschaft instabiler Tonigkeiten ist deren Potential, in verschiedener Weise umgedeutet oder auch alteriert werden zu können. Dies wiederum hat zur Folge, dass auch der akustische Klang des Dominantseptakkordes in unterschiedlichen Formen gedeutet werden und als vagierender Akkord auch in ganz unterschiedlichen Tonarten zum Vorschein kommen kann. Abbildung 5.15 stellt vier verschiedene *Umdeutungsmöglichkeiten* des Klanges *G-h-d-F* im tonartbezogenen Terzkreis gegenüber, ergänzt durch weiterführende Literaturangaben. Wie in Abbildung 5.15 gesehen werden kann, besitzen die Tonigkeiten des Dominantseptakkordes das Potential, auf ganz unterschiedliche tonale Zentren zu zielen. Der Dominantseptakkord ist damit ein gutes Modulationsmittel.

²³ Zum einen strebt der Durakkord mit hinzugefügter kleiner Septime als Dominantseptakkord in Richtung der leitereigenen Tonika, zum anderen kann er als übermäßiger Quintsextakkord in das im Quintenzirkel gegenüberliegende tonale Zentrum streben [83, S. 50].

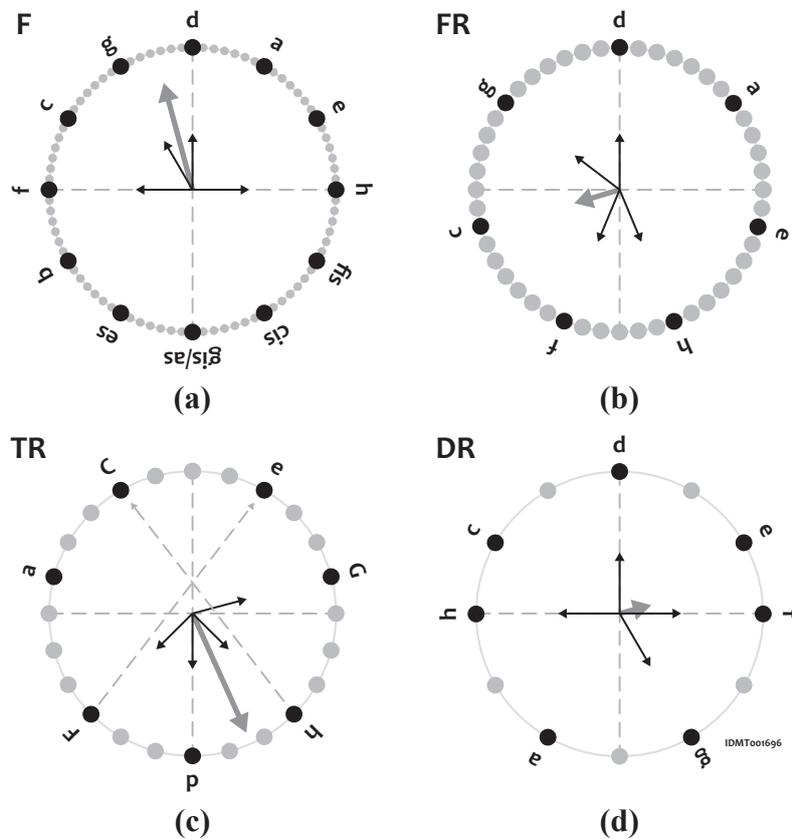


Abbildung 5.14 – Repräsentation des Dominantseptakkordes im Symmetriemodell: (a) im tonartübergreifenden Quintenzirkel *F*, (b) im tonartbezogenen Quintenzirkel *FR*, (c) im tonartbezogenen Terzkreis *TR* und (d) im tonartbezogenen diatonischen Tonleiterkreis *TR*

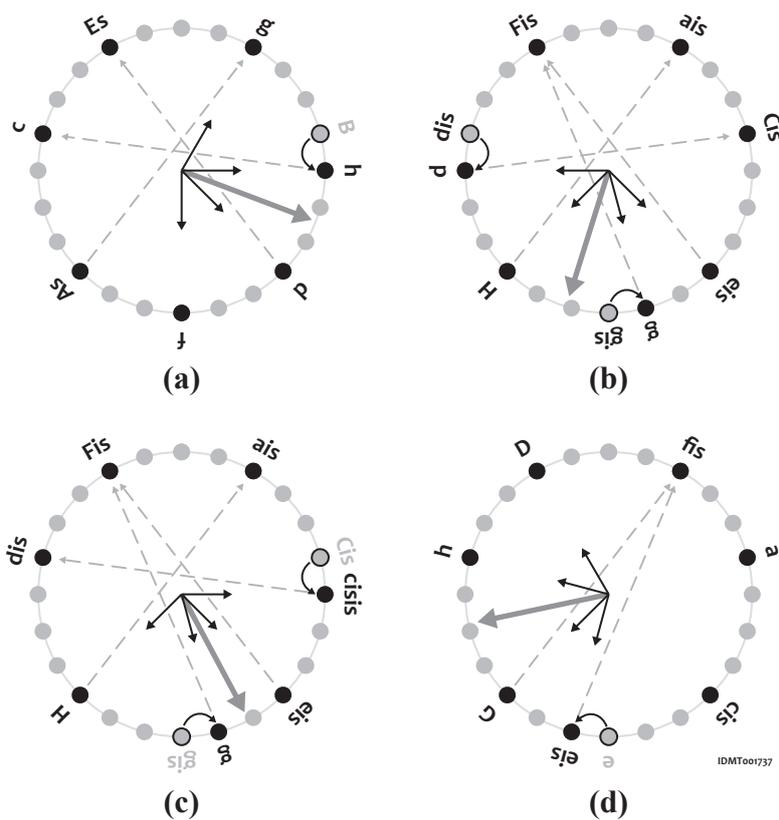


Abbildung 5.15 – Umdeutungsmöglichkeiten des Klanges *G-h-d-F*: (a) als Dominantseptakkord in *c*-Moll, (b) als übermäßiger Sextakkord („verminderter Septakkord auf der tiefalterierten Quinte“) in *Fis*-Dur [83, S. 48], (c) als verminderter Septakkord auf der tiefalterierten Dominantseptime in *dis*-Moll [83, S. 52] und (d) als Dursubdominante mit hochalterierter Sexte in *D*-Dur [83, S. 51]

5.4.3 Zwischendominanten

Zwischendominanten sind nicht leitereigene Klänge, die jedoch ein starkes Auflösungsbestreben auf einen leitereigenen Klang aufweisen. In der Regel handelt es sich bei Zwischendominanten um Durakkorde mit hinzugefügter kleiner Septime (siehe Kapitel Dominantseptakkord 5.4.2). Im tonartbezogenen Terzkreis lassen sich die Eigenschaften von Zwischendominanten gut darstellen. Wie in Abbildung 5.16 am Beispiel der Zwischendominanten C^7 (c, e, g, b) und A^7 (a, cis, e, g) veranschaulicht, belegen Zwischendominanten im *TR* vier benachbarte Tonigkeiten. Um einen *leitereigenen Durakkord* in eine Zwischendominante umzuwandeln, muss dessen große Septime erniedrigt werden. Dies ist in Abbildung 5.16a für den C-Dur-Akkord in Form des zu b erniedrigten h dargestellt. Durch die Erniedrigung entsteht eine Strebetonigkeit auf die im *TR* gegenüberliegende Tonigkeit a . Die Terz e des Dreiklanges verwandelt sich ebenfalls in eine Strebetonigkeit, welche auf die gegenüberliegende Tonigkeit f zielt. Es wird offensichtlich, dass der Akkord C^7 ein starkes Auflösungsbestreben in Richtung des Dreiklanges $F-a-C$ aufweist. Um einen *leitereigenen Mollakkord* in eine Zwischendominante zu verwandeln, muss dessen Terz erhöht werden. Im Beispiel des Dreiklanges a -Moll (Abbildung 5.16b) wird die Tonigkeit c nach cis erhöht, wodurch eine Strebetonigkeit auf das gegenüberliegende d entsteht. Die Terz cis und die Quinte e des so entstandenen A -Dur-Dreiklanges verwandelt sich in Strebetonigkeiten auf die gegenüberliegenden Tonigkeiten f und d . Eine starke Strebewirkung in Richtung des Dreiklanges $d-F-a$ entsteht.

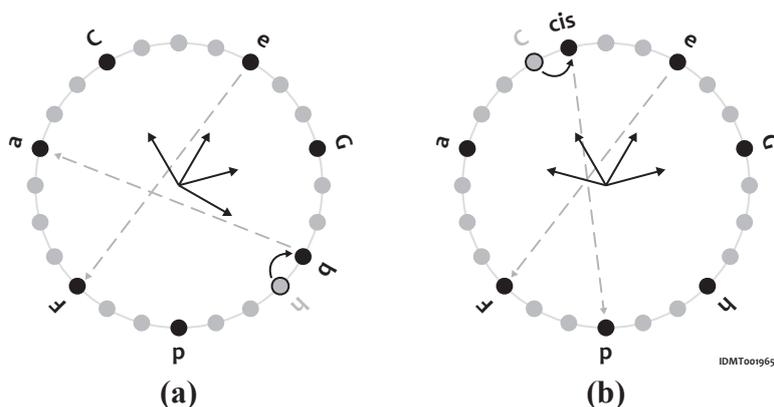


Abbildung 5.16 – Repräsentation von Zwischendominanten im TR: (a) C -Dur⁷ (b) D -Dur⁷

5.4.4 Der Sixte ajoutée

Der Sixte ajoutée [84, S. 115] wurde bereits im Zusammenhang mit charakteristischen Zusatzdissonanzen angesprochen (Abschnitt 5.3.3). Da sowohl der Dominantseptakkord

als auch der Sixte ajoutée im Rücken des TRs liegen und eine Reihe gemeinsamer Tonigkeiten besitzen, kommt es in beiden Fällen zu räumlich dicht beieinander liegenden Summenvektoren. Das deutlich stärkere Unterscheidungsmerkmal zwischen beiden Akkorden sind die darauffolgenden Akkorde: Während dem Sixte ajoutée der Regel nach die Dominante folgt, kommt nach dem Dominantseptakkord die Tonika. An dieser Stelle wird die Bedeutung des Zeitbezuges zur Unterscheidung von Akkorden deutlich.

5.4.5 Der verminderte Septakkord

Der verminderte Septakkord [84, S. 100] entsteht durch Aufeinanderschichten von vier Tönen im Abstand je einer kleinen Terz. Funktional wird dieser Akkord als ein kleiner Dominantnonenakkord ohne Grundton interpretiert. In seinem Buch „Harmonielehre“ [145] sagt Arnold Schönberg über den verminderten Septakkord folgendes aus (Siehe auch [84, S.105]): „Und es wird sich später noch zeigen, daß die Beziehungen, die dieser Akkord zu den Tonarten hat, noch viel reicher sind. Daß er in keiner eigentlich allein zu Hause, allein zuständig ist. Sondern daß er sozusagen überall heimatberechtigt und doch nirgends seßhaft ist, ein Kosmopolit oder ein Landstreicher! Ich nenne derartige Akkorde *vagierende Akkorde ...*“. Diese und weitere Eigenschaften des verminderten Septakkordes sollen nun im Kontext des Symmetriemodells besprochen werden:

- *Klanglich gesehen gibt es nur drei verschiedene verminderte Septakkorde:* Dies kann in Abbildung 5.17 anhand der drei unterschiedlich dargestellten Kreuze gesehen werden. Dadurch sind zwei verminderte Septakkorde von ihren Tonigkeiten her stets sowohl einen Halbton, einen Ganzton, eine große Terz als auch eine Quinte voneinander entfernt.
- *Der verminderte Septakkord kann keinem diatonischen Tonset eindeutig zugeordnet werden:* Im Symmetriemodell wird dies im tonartübergreifenden Quintenzirkel F deutlich: Hier spannen die Tonigkeitsvektoren des verminderten Septakkordes ein Kreuz auf, das alle Tonarten umspannt (Abbildung 5.17, solides Kreuz). Die Länge des *Summenvektors* des verminderten Septakkordes im tonartübergreifenden Quintenzirkel ist deshalb auch 0 und der Winkel somit undefiniert.
- *Der verminderte Septakkord ist in allen Tonarten zu Hause:* In Abbildung 5.17 wurden die drei verminderten Septakkorde in Form von Pfeilen und zugehörigen Funktionsbezeichnungen der diatonischen Tonart C-Dur zugeordnet: Der verminderte Septakkord *d-f-as-h* wirkt als Dominante, die sich in die Tonika auflösen muss, der verminderte Septakkord *fis-a-c-es* als Doppeldominante, die sich in die Dominante

auflösen muss, und der verminderte Septakkord *e-g-b-cis* als Zwischendominante, die sich in die Subdominantparallele auflösen muss.²⁴

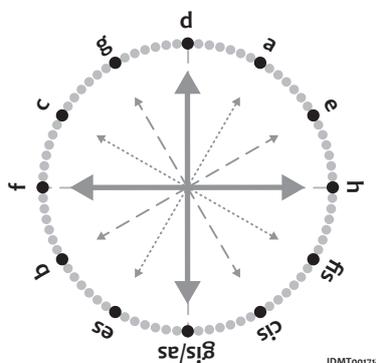


Abbildung 5.17 – Repräsentation des verminderten Septakkordes im tonartübergreifenden Quintenzirkel F: Dick durchgezogenes Kreuz: *h-d-f-as/gis*; gestricheltes Kreuz: *e-g-b-des/cis*; gepunktetes Kreuz: *a-c-es-ges/fis*

Aus Abbildung 5.17 wird darüber hinaus noch eine Besonderheit der Hierarchiedarstellung des Symmetriemodells deutlich. Hier wurde neben dem tonartübergreifenden Quintenzirkel auch noch der tonartübergreifende Terzkreis *T* eingezeichnet. Es ist zu sehen, dass die Ausläufer des den verminderten Septakkord bildenden Kreuzes im *F* auf die Stellen im *T* zeigen, wo sich die Schwerpunkte der Akkorde befinden, in die sich der verminderte Septakkord auflösen kann²⁵. Damit sind wir auch bereits bei der tonartbezogenen Betrachtung des verminderten Septakkordes. Wie bereits ausgeführt wurde, kann ein gegebener verminderte Septakkord vier diatonischen Tonsets und damit vier Durtonarten und vier Molltonarten zugeordnet werden.

In Abbildung 5.18 ist der verminderte Septakkord im tonartbezogenen Terzkreis einmal für C-Dur und einmal für *a*-Moll eingezeichnet. Zunächst kann man sehen, dass nur drei der vier Tonigkeiten im *TR* repräsentiert werden, nämlich die den verminderten Dreiklang *h-d-f* bildenden Tonigkeiten. Um die vierte Tonigkeit (*gis* oder *as*) in den tonartbezogenen Terzkreis einordnen zu können, muss zuvor das Tongeschlecht der aktuellen Tonart analysiert werden. Je nachdem ob es sich um eine Dur- oder Molltonart handelt, muss die Tonigkeit „*gis/as*“ entweder als tiefalteriertes *a* oder als hochalteriertes *g* in den *TR* eingeordnet werden (Abschnitt 5.3.6). Durch diesen Schritt kommt es zur

²⁴ An dieser Stelle soll bemerkt werden, dass sich die korrekte enharmonische Bezeichnung der Tonigkeiten automatisch aus dem tonartübergreifenden Quintenzirkel ergibt.

²⁵ So zeigt der in den Kreis F eingezeichnete verminderte Septakkord folgende Akkorde im tonartübergreifenden Terzkreis: *Es*-Dur und *c*-Moll, *C*-Dur und *a*-Moll, *A*-Dur und *fis*-Moll, *Fis*-Dur und *dis*-Moll.

Hinzufügung einer weiteren Strebetonigkeit (Abschnitt 5.3.5), welche im Fall von Moll auf den Tonikagrundton und im Fall von Dur auf die Tonikaquinte zeigt²⁶, so dass in beiden Fällen alle Tonikatonigkeiten (a, C und e) eine Strebetonigkeit besitzen. In Abbildung 5.18 ist dies durch entsprechende Pfeile visualisiert für C-Dur (Abbildung 5.18a) und a-Moll (Abbildung 5.18b) dargestellt.

Weiterhin ergibt sich durch die *korrekte* Einordnung der nichtdiatonischen Tonigkeit *gis/as* in den *TR* zum einen eine deutliche *Verlängerung des Summenvektors*, zum anderen *verändert* sich der Winkel des Summenvektors so, dass er weiter entfernt vom Summenvektor der zugehörigen Tonika befindlich ist. Sowohl die Verlängerung des Summenvektors als auch die weitere Entfernung vom Summenvektor der Tonika geben die in der Theorie beschriebene Erhöhung der dominantischen Wirkung des verminderten Septakkordes richtig wieder. Durch die richtige Einordnung der Tonigkeit „gis/as“ kann auch gleichzeitig der Grundton des Akkordes bestimmt werden, da dieser gegen den Uhrzeigersinn in direkter Benachbarung zum jeweiligen verminderten Septakkord liegt²⁷.

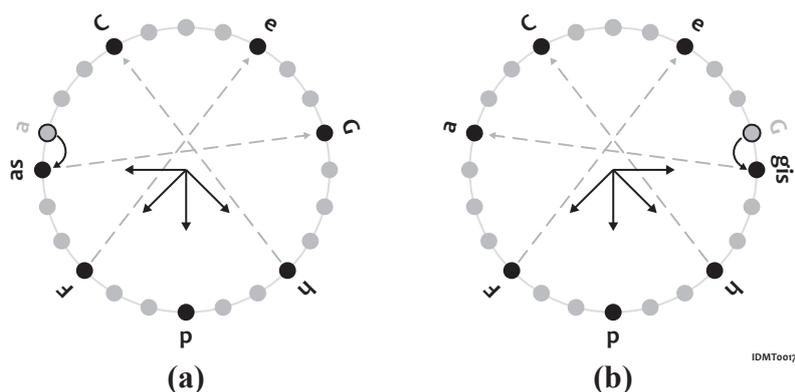


Abbildung 5.18 – Bei der Einordnung des verminderten Septakkordes in den *TR* muss zwischen Dur- (a) und Molltonarten (b) unterschieden werden.

5.4.6 Der übermäßige Dreiklang

Durch Hochalteration der Dominantquinte in Dur entsteht der übermäßige Dreiklang [83, S. 45]. Die hochalterierte Dominantquinte agiert als Strebetonigkeit in die Terz der Durto-

²⁶ Im Beispiel von Abbildung 5.18a (Dur) entsteht eine Strebetonigkeit „As“, welche auf die gegenüberliegende Tonikaquinte „G“ zielt. Im Beispiel von Abbildung 5.18b (Moll) entsteht eine Strebetonigkeit „Gis“, die auf den Tonikagrundton „a“ zielt.

²⁷ Im Fall von *gis-h-d-f* ist dies die Tonigkeit „E“, im Fall von „*h-d-f-as*“ ist dies die Tonigkeit „G“. Die hier dargestellte Gesetzmäßigkeit kann für die Entwicklung von Algorithmen zur korrekten Grundtonbestimmung verwendet werden.

nika (Abbildung 5.19b). Wie in Abbildung 5.19a gesehen werden kann, ist die Länge des *Summenvektors* des übermäßigen Dreiklangles im tonartübergreifenden Quintenzirkel F gleich 0, was zeigt, dass der übermäßige Dreiklang ohne weiteren Kontext nicht eindeutig einer Tonart zugeordnet werden kann. Der tonartübergreifende Charakter des übermäßigen Dreiklangles prädestiniert diesen zum Modulationsmittel (siehe Abschnitt 5.5.3).

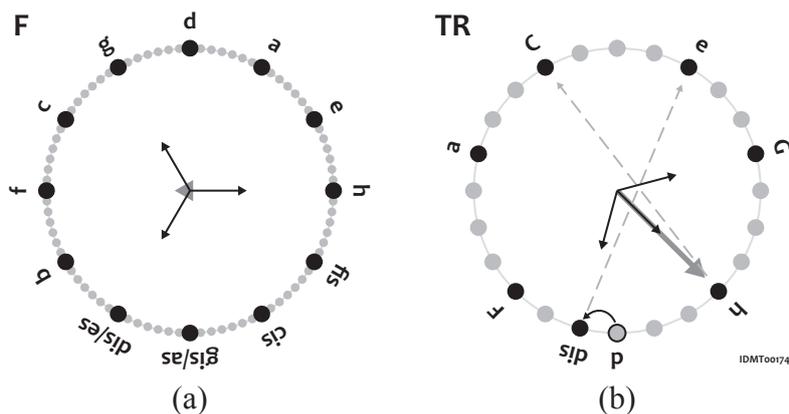


Abbildung 5.19 – Darstellung des übermäßigen Dreiklangles (a) im F und (b) im TR

5.5 Weitere Elemente

5.5.1 Kirchentonarten

Die Eigenschaften Kirchentonarten bzw. Modi (Abschnitt 2.6.4) werden im Symmetriemodell besonders durch den Summenvektor im tonartbezogenen Quintenzirkel FR veranschaulicht. Entsprechend der Dur-Moll-Separierung im FR (Abschnitt 5.3.1) befinden sich die Summenvektoren der Tonika von mollartigen Kirchentonarten (dorisch, äolisch und phrygisch) im FR rechts und von durartigen Kirchentonarten (lydisch, ionisch und mixolydisch) links von der Symmetrieachse (Abbildung 5.20). Die Summenvektoren der Tonikadreiklänge der am häufigsten vorkommenden Modi „Ionisch“ und „Äolisch“ belegen jeweils das zentrale Mittensegment auf beiden Hälften.

5.5.2 Harmonisch Moll

Die reine Molltonleiter besitzt keinen Leitton, d.h. es fehlt eine Strebetonigkeit in den so bedeutsamen Tonikagrundton²⁸. Dies bedeutet, dass die Dominante nur ein schwaches Auflösungsbestreben in Richtung der Tonika aufweist. Dieses Problem wird in der

²⁸ Im Beispiel der Tonart a -Moll-rein fehlt ein Leitton gis nach a .

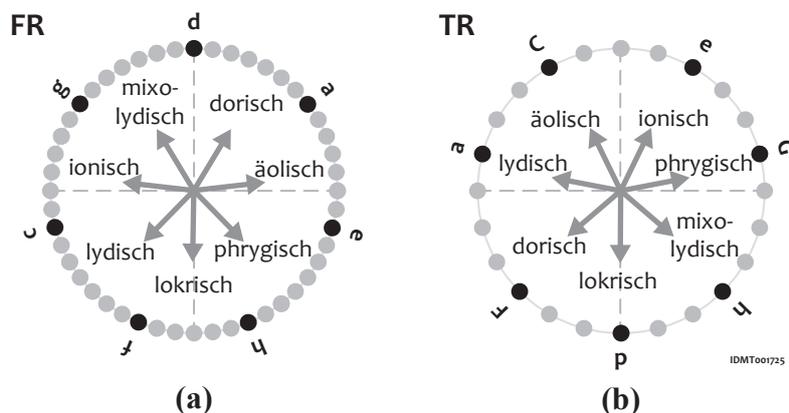


Abbildung 5.20 – Einordnung der Summenvektoren der Tonikadreiklänge der sieben Kirchentonarten in das Symmetriemodell: (a) *FR*, (b) *TR*

musikalischen Praxis durch Erhöhung der Dominantterz gelöst²⁹, wodurch eine auf den Tonikagrundton zielende Strebetonigkeit und damit ein Leitton entsteht (siehe auch Abbildung 5.12c). Durch diesen Schritt gleichen sich die strukturellen Eigenschaften von Dur und Moll an, teilweise übertrifft harmonisch Moll sogar die Qualität von Dur: Während Dur nur zwei Strebetonigkeiten (*f* und *h*) auf den Tonikagrundton (*C*) und die Tonikaterz (*e*) hat, besitzt harmonisch Moll drei Strebetonigkeiten (*gis*, *h* und *f*) auf alle drei Tonigkeiten der Tonika (*a*, *C*, *e*). Darüber hinaus liegt wie im Fall von Dur im *TR* durch die Erhöhung der Dominantterz dem Tonikagrundton und der Tonikaterz ein verminderter Dreiklang gegenüber. Die Dominante enthält wie im Fall von Dur den Leitton und weist dadurch ein starkes Auflösungsbestreben in Richtung der Tonika auf.

5.5.3 Ausweichungen und Modulationen

Unter dem Begriff der Modulation versteht man den musikalischen Übergang von einer Tonart in eine andere [83, S. 63]. Je nachdem wie die Modulation gestaltet ist, kann es hierbei zu einem fließenden, kaum wahrnehmbaren Wechsel in die neue Tonika oder auch zu einem deutlichen „tonalen Ruck“ kommen. In der Musikgeschichte haben sich verschiedene Modulationstypen herausgebildet, deren Einsatz ein wichtiges Genremerkmal darstellen könnte. Die Ausführung dieser einzelnen Typen würde an dieser Stelle über den Fokus der vorliegenden Arbeit hinausgehen. Die meisten Modulationsformen sind jedoch dadurch gekennzeichnet, dass zwischen der Ausgangs- und der Zieltonart tonales Material gespielt wird, das beiden Tonarten gemeinsam ist bzw. welches beide Tonarten

²⁹ Im Beispiel von *a*-Moll wird die Tonigkeit *G* auf *gis* erhöht. Dadurch entsteht eine Strebetonigkeit in Richtung des Tonikagrundtones *a*.

verbindet. Solche Verbindungsmöglichkeiten, als Modulationsmittel bezeichnet, sollen im Folgenden kurz beschrieben werden:

- Soll in quintbenachbarte oder parallele Tonarten moduliert werden, dann können bereits leitereigene Dreiklänge die verbindende Schnittstelle sein. In diesem Fall reicht es aus, die Funktion oder das Tongeschlecht leitereigener Klänge umzudeuten.
- Eine weitere Möglichkeit ist der Einsatz von Zwischendominanten (Abschnitt 5.4.3): Diese bieten zum einen starke *Strebetonigkeiten auf Klänge der alten Tonart*, zum anderen können sie als Dominantseptakkord der neuen Tonart (Abschnitt 5.4.2) umgedeutet werden. Dem Komponisten bleibt damit die Möglichkeit, sowohl in der alten Tonart fortzufahren als auch in die neue Tonart zu wechseln.
- Eine sehr häufig verwendete Möglichkeit ist der Einsatz des verminderten Septakkordes. Wie in Abschnitt 5.4.5 gezeigt wurde, besitzt er als Dominante, Doppeldominante oder Zwischendominante Bedeutung in allen Tonarten und kann deshalb auch als Verbindungsglied zwischen allen Tonarten agieren. Ähnliches gilt für den übermäßigen Dreiklang, der – wie der verminderte Septakkord – alle Tonarten miteinander verbindet (Abschnitt 5.4.6).
- Um größere tonale Distanzen zu überwinden, bietet sich auch die in Abschnitt 5.4.2 skizzierte Umdeutung des Dominantseptakkordes in einen übermäßigen Quintsextakkord an [83, S. 79]. Wie in Abbildung 5.15a und 5.15c am Beispiel der Tonarten „C-Dur“ und „Fis-Dur“ veranschaulicht, können damit im Quintenzirkel gegenüberliegende tonale Zentren verbunden werden.
- In vielen Fällen werden auch Kombinationen der bisher besprochenen Möglichkeiten eingesetzt. Dadurch werden die tonalen Bewegungsmöglichkeiten noch einmal deutlich erweitert.

Weiterführende Arbeiten sollten untersuchen, ob es möglich ist, Modulationstypen anhand der Bewegungspfade des Summenvektors im Tonraum zu klassifizieren. Die oben genannten Punkte zeigen, dass Modulationen in vielen Fällen Gesetzmäßigkeiten folgen, die im Symmetriemodell sehr einfach beschrieben und entsprechend in Algorithmen modelliert werden können.

5.5.4 Stabilität von Tönen in einem gegebenen Kontext

In Abschnitt 2.7.1 wurden Carol L. Krumhansls Testtonprofile³⁰ vorgestellt (Abbildung 2.5). Diese beschreiben die Stabilität von Tonigkeiten in einem gegebenen tonalen Kontext. Ein Ziel des Symmetriemodells besteht darin, diese Stabilitäten vorherzusagen. Die zugrundeliegenden Annahmen hierbei bestehen darin, dass erstens die gewählte Tonart durch den Grundton repräsentiert wird und dass zweitens sich die Stabilität einer beliebigen Tonigkeit innerhalb dieser Tonart in der geometrischen Beziehung zwischen Tonigkeit und Grundton widerspiegelt. Ob dies funktioniert bzw. ob ein Zusammenhang zwischen den Distanzen im Symmetriemodell und den von Carol Krumhansl empirisch ermittelten Stabilitätsprofilen besteht, soll im Folgenden am Beispiel der Stabilitäten von Tonigkeiten in der Tonart C-Dur diskutiert werden. Es ist zunächst anzumerken, dass die folgenden Betrachtungen nur die diatonischen Tonigkeiten c, d, e, f, g, a und h berücksichtigen. Die restlichen Tonigkeiten fis, cis, gis, dis und ais unterscheiden sich in ihrer Stabilität nicht sehr und können durch einen konstanten Wert angenähert werden. Um eine bessere Vergleichbarkeit zu gewährleisten, werden weiterhin Krumhansls und alle im Symmetriemodell berechneten Profile auf einen Wertebereich zwischen 0,0 und 1,0 skaliert. Es wird nun die Nähe der diatonischen Tonigkeiten c, d, e, f, g, a und h zur Tonigkeit c berechnet, da diese den Grundton der gewählten Tonart repräsentiert. Die Berechnung erfolgte für die Kreise F, FR, TR und DR unabhängig. Eine Winkeldistanz von 0° entspricht dabei einer Nähe von 1,0, der Winkel der vom Grundton am weitesten entfernten Tonigkeit einer Nähe von 0,0. Das Ergebnis dieser Schritte sind vier verschiedene Profile, welche die Nähe der Tonigkeiten c, d, e, f, g, a und h zur Tonigkeit c in den Kreisen F, FR, TR und DR beschreiben. Diese Profile werden als P_F, P_{FR}, P_{TR} und P_{DR} bezeichnet und sind in Tabelle A.1 zu finden. Der nächste Schritt besteht nun darin zu überprüfen, inwiefern die einzelnen Profile mit dem von Carol L. Krumhansl für die Tonart C-Dur ermittelten Profil P_{KHL} übereinstimmen. Dazu werden die Pearson-Korrelationen [53] zwischen den Profilen $P_{F,FR,TR,DR}$ und dem Profil P_{KHL} berechnet.³¹ Die anschließende Analyse der Ergebnisse ergibt, dass die im DR berechnete Nähe (P_{DR}) in keinem oder nur geringem Zusammenhang mit der wahrgenommenen Stabilität von Tonigkeiten steht. Die Korrelation zwischen P_{DR} und P_K beträgt hier $r = 0,08$. Die Korrelationen der restlichen Profile $P_{F,FR,TR}$ mit dem Profil P_{KHL} liegen alle im Bereich um 0,7 und sind damit eher mäßig. Die sich nun ergebende Frage ist, ob sich durch eine Kombination der Profile $P_{F,FR}$ mit dem

³⁰ Als Stabilitätsprofil wird hier eine Reihe von zwölf Zahlenwerten bezeichnet, die den zwölf chromatischen Tonigkeiten $c, cis, d \dots$ je einen Stabilitätswert zuordnen.

³¹ Die Zwischenschritte sowie die Ergebnisse sind im Anhang (Tabelle A.1) zu finden.

Profil P_{TR} eine bessere Übereinstimmung mit dem Profil P_{KHL} ergibt. Hierbei konnten folgende Werte erzielt werden: Die Korrelation zwischen den Profilen P_{KHL} und P_{F+TR} beträgt $r=0,88$, die Korrelation zwischen P_{KHL} und P_{FR+TR} ist mit $r = 0,93$ etwas höher. Dies ist ein Hinweis darauf, dass der tonartbezogene Quintenzirkel FR tonale Stabilitäten etwas besser modelliert. Die Profile P_{KHL} und P_{FR+TR} sind in 5.21a zu sehen.

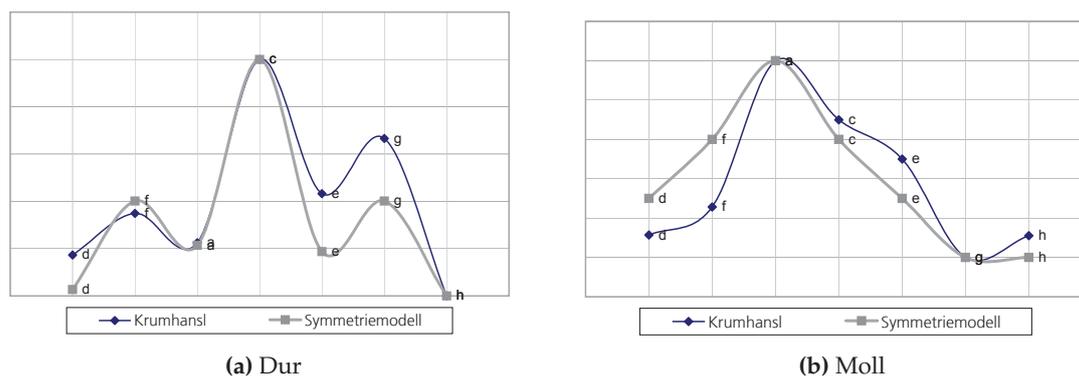


Abbildung 5.21 – Vergleich von Carol Krumhansls empirisch ermittelten Tonstabilitäten (Abschnitt 2.7.1) mit den im Symmetriemodell ermittelten Abstands-Profilen

Die zuvor für die Tonart C-Dur beschriebene Prozedur wurde auch für die Tonart a -Moll durchgeführt. Dazu wurden jedoch nicht die Nähe bezogen auf die Tonigkeit c , sondern auf die Tonigkeit a berechnet und mit Carol L. Krumhansls a -Moll-Profil verglichen. Hierbei ergab bereits die Korrelation der Profile P_{TR} und P_{KHL} einen relativ hohen Wert von $0,87$.³² Eine Kombination der Profile P_{TR} und P_F bzw. P_{FR} brachte keine Erhöhung der Korrelation. Die schrittweise Berechnung der Ergebnisse für a -Moll ist in Tabelle A.2 zu sehen, das resultierende Profil P_{TR} in Abbildung 5.22b.

Zusammenfassend lässt sich sagen, dass die Stabilitäten von Tonigkeiten innerhalb von Durtonarten durch eine Kombination der Distanzen im FR und TR am besten angenähert werden, die Stabilitäten von Tonigkeiten innerhalb von Molltonarten dagegen allein durch den TR . Dies ist ein Hinweis, dass je nach Kontext unterschiedliche tonale Gesetzmäßigkeiten berücksichtigt werden müssen. Eine weitere Erhöhung der Korrelation zwischen Krumhanslprofilen und Symmetriemodell könnte durch Einführung eines Leittonfaktors sowie einer höheren Stabilität für die ersten Obertöne des Grundtones erreicht werden. In Abbildung 5.21 ist zu sehen, dass die Töne h bei C-Dur und g bei a -Moll zu hohe Stabilitätswerte haben. Die ersten Obertöne des Grundtones (g und e bei C-Dur sowie e bei a -Moll) haben dagegen eine zu geringe Stabilität.

³² Laut [53] entspricht dies bei einem einseitigen Test einem Signifikanzlevel von $0,005$. Die Wahrscheinlichkeit eines zufälligen Zusammenhanges ist damit $0,5$ Prozent.

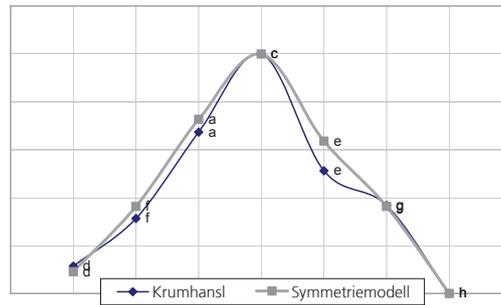


Abbildung 5.22 – Vergleich von Carol Krumhansls Testtonprofilen (Abschnitt 2.7.1) mit den Vorhersagen des Symmetriemodells: Gemittelte Dur- und Moll-Testtonprofile

Eine interessante Ergänzung ergibt sich noch für eine Betrachtung eines für C-Dur- und *a*-Moll gemittelten Profils (Abbildung 5.22). Dieses korreliert mit einem Faktor von 0,99 mit dem Profil P_{TR} für C-Dur. Dies zeigt, dass die geometrischen Abstände im *TR* die gemittelten Stabilitäten von Dur- und paralleler Molltonart perfekt repräsentieren.

Kapitel 6

Tonraumbasierte Synthese musikalischer Audiosignale

Eine gute Organisation von Tönen innerhalb von geometrischen Tonräumen erlaubt nicht nur Analyse und Vergleich von musikalischen Audiosignalen. Vielmehr ist es auch möglich, auf Basis der in dieser Arbeit entwickelten Erkenntnisse neue Möglichkeiten zum Erzeugen von Musik zu entwickeln. Während bei herkömmlichen Synthesizern hauptsächlich zeitliche und spektrale Eigenschaften der Klangfarbe gesteuert werden, liegt die Hauptaufmerksamkeit bei der tonraumbasierten Klangsynthese auf der Parametrisierung und einfachen Steuerung des zeitlich-sequentiellen oder simultanen Zusammenklings unterschiedlicher Töne. Darüber hinaus motivieren Tonräume, über neuartige Mensch-Maschine-Schnittstellen nachzudenken. Vorstellbar sind hierbei Szenarien, in denen ein Spieler sich durch einen Tonraum bewegt und die dort befindlichen Töne anfassen, auswählen und anhören kann. Die musikalische Logik des Tonraumes wird somit direkt erfahrbar, hörbar und erlebbar. Je nach Qualität der Interaktionsmechanismen kann sich der bisher so starre und theoretische Tonraum zu einem mächtigen Musikinstrument sowie Kompositions- und Improvisationswerkzeug entwickeln. Möglichkeiten in diese Richtung aufzuzeigen ist Gegenstand dieses Kapitels.

6.1 Motivation

Welchen Vorteil kann das Wissen um Tonräume für die Entwicklung von Musikinstrumenten bieten? Wie können Tonräume gewinnbringend für das Instrumentalspiel, das Komponieren oder Improvisieren eingesetzt werden? Um eine Antwort auf diese Fragen zu geben, soll zunächst ein einfaches Modell zur musikalischen Tonvorstellung und Interaktion herangezogen werden, welches durch Hugo Riemanns „Ideen zu einer ‘Lehre

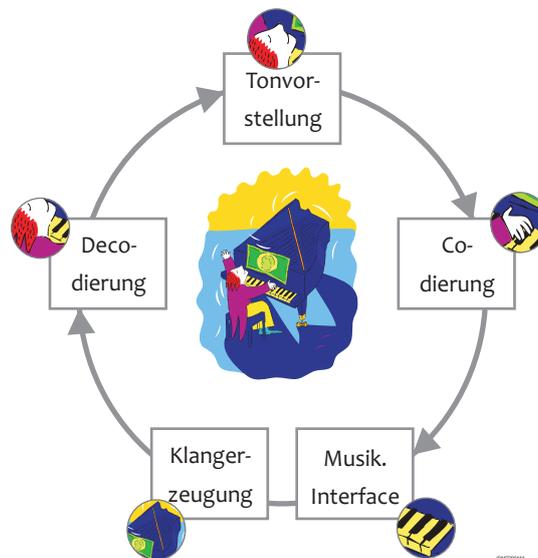


Abbildung 6.1 – Einfaches Modell zur Veranschaulichung musikalischer Tonvorstellung und Interaktion

von den Tonvorstellungen“ [133] inspiriert ist. Ausgangspunkt dieses in Abbildung 6.1 dargestellten Modells ist die *Tonvorstellung*: Dies können konkrete Melodien oder Akkordfolgen, aber auch bestimmte Gefühle oder Stimmungen sein. Um die Tonvorstellung hörbare Realität werden zu lassen, muss diese durch den Spielenden auf eine *musikalische Schnittstelle codiert* werden. Die musikalische Schnittstelle repräsentiert die Bedieneinheit eines Musikinstrumentes, was z.B. die Klaviatur eines Flügels sein kann. Die musikalische Schnittstelle ist wiederum mit einer *Klangerzeugung* verbunden, welche für die Abstrahlung eines Schallsignals sorgt. Dieses wird vom auditorischen System des Spielenden aufgenommen und wieder in eine Tonvorstellung zurück *decodiert*. Hierbei kommt es zu einem Abgleich zwischen erwarteter und tatsächlicher Tonvorstellung. Sind diese unterschiedlich, so muss der Spieler durch weiteres Üben die Codierung solange anpassen, bis eine Übereinstimmung zwischen ursprünglicher und tatsächlicher Tonvorstellung erreicht ist. Bei den meisten herkömmlichen Musikinstrumenten dauert dieser Prozess viele Jahre und erfordert intensives Üben. Ursache hierfür liegt vor allem darin, dass *Tonvorstellung und musikalische Schnittstelle oft nur wenig übereinstimmen* und deshalb komplizierte „Codierungsvorschriften“ erlernt werden müssen. Es ist deshalb erforderlich zu überlegen, inwiefern musikalische Schnittstellen so umgestaltet werden können, dass sie der menschlichen Tonvorstellung näher kommen. Der Vorschlag dieser Arbeit besteht deshalb darin, die Eigenschaften von Tonräumen auf musikalische Schnittstellen anzuwenden. Dies bedeutet, dass man Musikinstrumente entwickelt, deren Bedienelemente so angeordnet sind, dass sie der Anordnung von Tönen in einem gewählten Tonraum

entsprechen. Je stärker dieser wiederum die *tatsächlich wahrgenommenen musikalischen Eigenschaften widerspiegelt*, desto mehr entsprechen die musikalische Benutzerschnittstelle und damit die vom Spieler durchzuführenden motorischen Gesten der Tonvorstellung. Dies führt schließlich zu einfacheren und leichter erlernbaren Codierungsvorschriften. Im (sicherlich nicht erreichbaren) Idealfall würden Tonvorstellung und musikalische Schnittstelle korrespondieren und keine Codierung mehr erforderlich sein.

6.2 Aufbau eines Tonraumsynthesizers

Der Kern eines Tonraumsynthesizers ist der Tonraum, wie der Name schon sagt. Dies kann einer der Tonräume aus dem Stand der Technik (Kapitel 3) oder auch einer der Teilräume des Symmetriemodells sein (Abschnitt 4.2). Um die im Tonraummodell organisierten Töne tatsächlich zum Erklingen zu bringen, ist die Einführung der sogenannten *Tonselektionsfunktion* erforderlich. Diese beschreibt die Auswahl der zu spielenden Töne im Tonraum und wird durch den Spieler gesteuert. Die dritte Komponente ist die *Benutzerschnittstelle*, mit deren Hilfe der Spieler den Tonraum auswählt und konfiguriert und des Weiteren die Parameter der Tonselektionsfunktion steuert. Die vierte Komponente ist die Klangerzeugung, welche die selektierten Töne in einer gewählten Klangfarbe wiedergibt. Die Klangerzeugung ist nicht Gegenstand der vorliegenden Betrachtungen. Die einzelnen Komponenten eines Tonraumsynthesizers werden nun hier detailliert vorgestellt.

6.3 Der Tonraum

Der Stand der Technik (Kapitel 3) und auch das Symmetriemodell (Kapitel 4) stellen eine Reihe von Tonräumen bereit. Für einen Tonraumsynthesizer sind hierbei besonders diejenigen Tonräume interessant, welche häufig benötigte Tonkombinationen und Akkordfolgen geometrisch einfach (in Benachbarung, möglichst ein- bis zweidimensional) repräsentieren und des Weiteren möglichst die wahrgenommenen Tonbeziehungen wiedergeben. Je nach Art des zu erzeugenden musikalischen Ereignisses (Melodiefolge, Akkordfolge, Tonartwechsel etc.) sind ganz unterschiedliche Tonräume geeignet. So könnte sich z.B. das *Vial/Weber/Schönberg-Netz* (Abschnitt 3.2) als Benutzerschnittstelle eignen, um gezielt zwischen Tonarten zu wechseln. Der diatonische Tonleiterkreis *DR* des Symmetriemodells (Abschnitt 4.2.6) erlaubt mit seiner tonhöhenorientierten diatonischen Tonanordnung und mit sehr einfachen geometrischen Pfaden melodische Sequenzen zu beschreiben. Der *tonartbezogene Terzkreis TR* des Symmetriemodells hebt besonders die Gruppierung von Tönen zu Dur- und Molldreiklängen hervor (Kapitel 5) und unterstützt

dadurch ein vereinfachtes Akkordspiel. Der *tonartübergreifende Quintenzirkel F* veranschaulicht die Zirkularität der Tonarten unter Annahme einer temperierten Stimmung und ist damit eine Alternative zum Vial/Weber/Schönbergnetz. Die bis hierher genannte Liste von Möglichkeiten könnte noch lange erweitert werden.

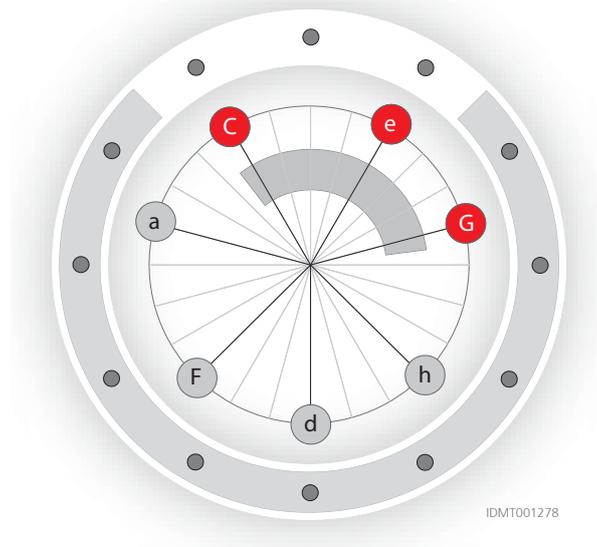


Abbildung 6.2 – Das HarmonyPad

Um das Thema „tonraumbasierte Klangsynthese“ möglichst anschaulich und plastisch zu gestalten, sollen die nachfolgenden Ausführungen an einem konkreten tonraumbasierten Musikinstrument erläutert werden, nämlich dem HarmonyPad. Das HarmonyPad ist in Abbildung 6.2 dargestellt. Wie hier gesehen werden kann, kombiniert das HarmonyPad zwei Tonräume: Der innere Tonraum ist der in Abschnitt 4.2.5 beschriebene *tonartbezogene Terzkreis TR*. Dieser Kreis ordnet Töne so an, dass benachbarte Tonigkeiten stets wohlklingende Dur- und Molldreiklänge bilden. Durch die Auswahl von nur sieben diatonischen Tonigkeiten ist die Komplexität des Kreises sehr gering. Das HarmonyPad kann deshalb vom Spieler schnell verstanden werden, was zu wichtigen Erfolgserlebnissen gerade für Anfänger führt. Die geometrische Zuordnung von charakteristischen Zusatzdissonanzen (Abschnitt 5.3.3) ermöglicht darüber hinaus, auch schon sehr bald wichtige Akkorde wie den Sixte Ajoutée (Abschnitt 5.4.4) oder den Dominantseptakkord (Abschnitt 5.4.2) spielen zu können. Die Visualisierung des tonalen Zentrums, die Zuordnung von Funktionen (Abschnitt 5.3.1), die Darstellung von Strebetonigkeiten (Abschnitt 5.3.5) oder auch die Möglichkeit, Tonigkeiten sinnvoll zu alterieren (Abschnitt 5.3.6), eröffnen darüber hinaus auch dem fortgeschrittenen Musiker viele kreative Möglichkeiten und einen tieferen Einblick in die musikalische Logik. Da der tonartbezogene Terzkreis jedoch nur die sie-

ben Tonigkeiten eines bestimmten diatonischen Tonsets darstellt, ist eine Möglichkeit erforderlich, diesen für andere Tonarten umzukonfigurieren. Dies erfolgt mit Hilfe des in Abbildung 6.2 im äußeren Ring dargestellten Tonraumes, dem *tonartübergreifenden Quintenzirkel F* (Abschnitt 4.2.2). Jeder der darin dargestellten zwölf grauen Punkte repräsentiert ein diatonisches Tonset. Durch Anwahl der entsprechenden Tonart wird der im inneren dargestellte *TR* auf das entsprechende Tonset umkonfiguriert.

6.3.1 Modellierung von Tonhöhe

Das Symmetriemodell, wie es bisher vorgestellt wurde, ist ein reines Tonigkeitsmodell (Abschnitt 2.3.3). Für die Analyse von Musiksignalen ist dies für viele Anwendungsfälle völlig ausreichend. Die tonraumbasierte Synthese von Musiksignalen dagegen kann mit der Tonigkeit allein nichts anfangen. Erst wenn die Oktavlage (n) einer Tonigkeit bekannt ist (sind), kann ein hörbares Musiksignal erzeugt werden. Um beim Beispiel aus Abbildung 6.2 zu bleiben: Es ist erforderlich, eine Möglichkeit zu finden, die unterschiedlichen Oktavlagen der sieben im *TR* dargestellten Tonigkeiten zu repräsentieren. In Anlehnung an Wilhelm Moritz Drobischs Spiralmodell besteht eine mögliche Lösung darin, den zweidimensionalen *TR* um eine dritte, die Tonhöhe repräsentierende Dimension zu erweitern. Ergebnis wäre ein dreidimensionales spiralförmiges Modell ähnlich dem in Abbildung 2.2a dargestellten. Problematisch bei dieser Lösung ist jedoch, dass durch das Hinzukommen einer dritten Dimension die Auswahl von Tönen und damit die Interaktion im Tonraum deutlich komplizierter wird. Eine *bessere Lösung* besteht deshalb darin, den *TR* zunächst einmal in einem Polarkoordinatensystem zu betrachten. Hier stellt man fest, dass die Tonigkeiten im *TR* nur die Winkeldimension belegen, die radiale Dimension ist noch frei und kann zur Repräsentation der Tonhöhe verwendet werden. Die sich dadurch ergebende Tonanordnung ist in Abbildung 6.3 dargestellt: Die Winkeldimension des tonartbezogenen Terzkreises wurde auf der als *Tonigkeitsachse* bezeichneten Ordinate eines 2D-Koordinatensystems abgebildet. Die radiale Dimension des TRs wurde auf der als *Tonhöhenachse* bezeichneten Abszisse abgebildet. Die *konkrete geometrische Position* von Tönen (Abbildung 6.3, schwarze Punkte) ergibt sich aus der Zugehörigkeit von Tonhöhen zu Tonigkeiten.

Je nach Wahl des Tonraumes kann die Reihenfolge der Tonigkeiten auf der Tonigkeitsachse variieren. Die zweidimensionale Anordnung von Tönen bezogen auf Tonhöhe und Tonigkeit bezeichnet der Autor allgemein als *Tonigkeits-Tonhöhenraum*. Dieser ist von grundlegender Bedeutung, da er unabhängig von der gewählten Tonigkeitsanordnung sinnvolle musikalische Akkordtransformationsoperationen ermöglicht. Dazu ein Beispiel: In Abbildung 6.3 wurde ein grau hinterlegtes Rechteck eingezeichnet, welches

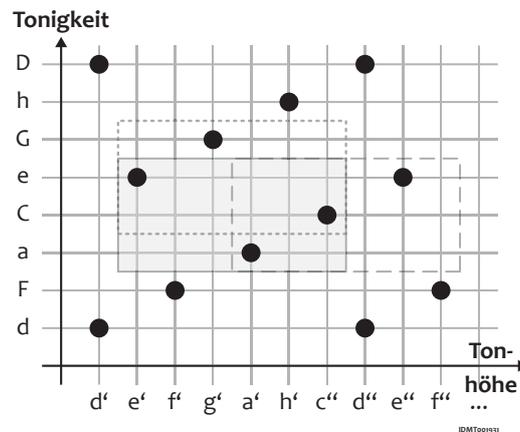


Abbildung 6.3 – Der Tonigkeits-Tonhöhenraum am Beispiel des TRs

so dimensioniert ist, dass es die Töne e' , a' und c'' überdeckt. Man stelle sich vor, dass die überdeckten Töne tatsächlich erklingen. Verschiebt man das Rechteck nun in horizontaler Richtung entlang der Tonhöhenachse (Abbildung 6.3, gestrichelter Rahmen), so fällt der Ton e' aus der Auswahl heraus und der Ton e'' kommt dafür in die Auswahl hinein, der ursprüngliche Akkord wurde durch eine einfache Translationsoperation in eine *andere Umkehrung* transformiert. Verschiebt man das graue Rechteck in vertikaler Richtung, also entlang der Tonigkeitsachse, so entfällt der Ton a' aus der Auswahl und der Ton g' kommt dafür hinzu. Der ursprüngliche a -Moll-Akkord wurde in einen C-Dur-Akkord transformiert. Die Transformation wird dabei so durchgeführt, dass es *automatisch zu einer günstigen Akkordverbindung bzw. optimalen Stimmenführung* kommt. Dies und auch die Vermeidung von Quintparallelen vereinfachen das Spiel des Tonraumsynthesizers wesentlich und stellen ein wichtiges Klangqualitätsmerkmal dar.

6.4 Die Tonselektionsfunktion

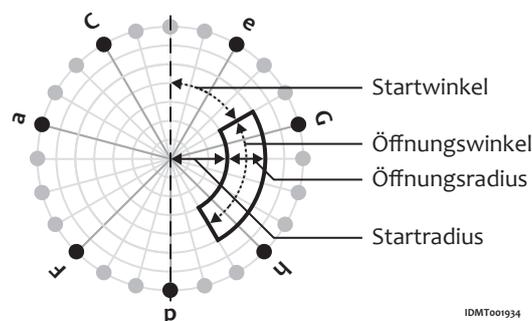


Abbildung 6.4 – Die Parameter der Tonselektionsfunktion

Nachdem mit Hilfe des Tonraumes Töne geometrisch bedeutungsvoll angeordnet wurden, muss nun für einen Mechanismus gesorgt werden, der es erlaubt, diese Töne gezielt zum Klingen zu bringen. Dieser Prozess wird als *Tonselektion* bezeichnet und durch die sogenannte *Tonselektionsfunktion* beschrieben. Im Fall zweidimensionaler Tonräume wie dem *TR* definiert diese Funktion einen als *Selektionsfläche* bezeichneten Raumausschnitt. Das Ziel dieses Abschnittes besteht darin, ein möglichst kleines und einfaches Parameter set zu bestimmen, das es dem Spieler des Tonraumsynthesizers erlaubt, solche Selektionsflächen zu erzeugen und so gewünschte Tonkombinationen und Tonfolgen schnell und effizient umzusetzen. Der Autor schlägt hierzu vor, die Tonauswahl über vier Parameter zu steuern, nämlich den Startwinkel, den Startradius, den Öffnungswinkel und den Öffnungsradius (Abbildung 6.4).

Der *Startwinkel* definiert den Winkel, an welchem der Auswahlbereich beginnt. Abbildung 6.5 zeigt den Auswahlbereich mit drei verschiedenen Startwinkeln: Der Öffnungswinkel ist so eingestellt, dass drei benachbarte Tonigkeiten überdeckt werden. Dadurch ist es möglich, eine komplette Akkordkadenz durch alleiniges Verändern des Startwinkels zu erzeugen. Bei der in Abbildung 6.5 dargestellten Akkordkadenz handelt es sich um die Kadenz *F-Dur*, *C-Dur* und *G-Dur*. Die aktiven Töne sind dunkel markiert. Weiterhin ist es möglich, von einem Akkord in einen anderen überzublenden, indem man den Auswahlbereich von einem Punkt zu einem anderen Punkt stufenlos verschiebt.

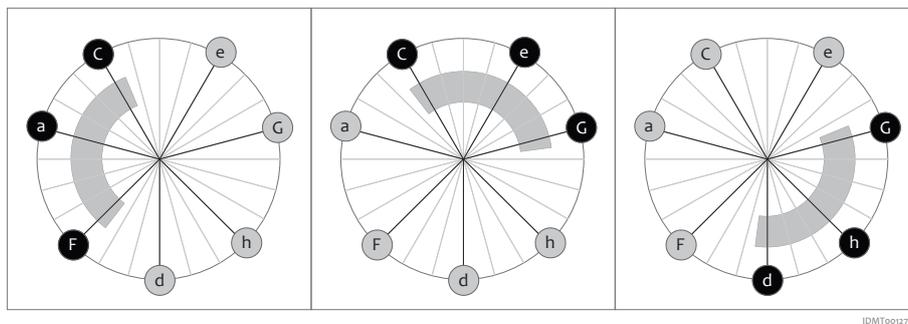


Abbildung 6.5 – Erzeugen einer Akkordkadenz durch Verändern des Startwinkels

Durch den *Öffnungswinkel* wird festgelegt, wie viele Töne durch den Auswahlbereich erfasst werden. Ein kleiner Öffnungswinkel – wie in Abbildung 6.6a gezeigt – umfasst nur eine einzige Tonigkeit. Durch kontinuierliches Vergrößern des Öffnungswinkels kann ein einzelner Ton in ein Terzintervall und anschließend in einen kompletten Akkord bis hin zu einem Vierklang überblendet werden (Abbildung 6.6). Dadurch ist es möglich, Einzel- oder Zweitonarpeggios oder auch melodische Linien zu spielen. Im Beispiel in Abbildung 6.6 wird der Ton C in das Intervall *C-e* und anschließend den Dur-Dreiklang *C-e-G* überführt.

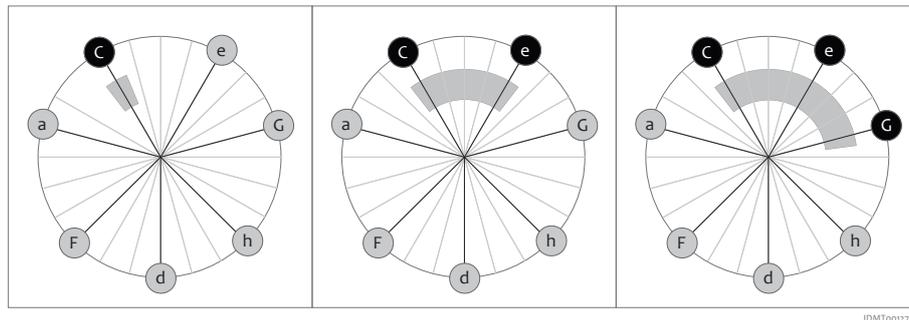


Abbildung 6.6 – Der Öffnungswinkel definiert die Anzahl der gleichzeitig hörbaren Töne

Der *Startradius* definiert die Oktavlage des erklingenden Akkordes. Entsprechend Abschnitt 6.3.1 erzeugt eine Bewegung des Auswahlbereiches in radialer Richtung unterschiedliche Umkehrungen. Die Tonhöhe steigt dabei mit Erhöhung des Radius an. Im Beispiel aus Abbildung 6.7a ist zunächst ein C-Dur-Akkord in tieferer Oktavlage ausgewählt. Anschließend wird der Auswahlbereich entlang der radialen Dimension, d.h. entlang der Tonhöhenachse (Abschnitt 6.3.1), bewegt. Durch diesen Schritt wird der ausgewählte C-Dur-Akkord in höhere Oktavlagen verschoben, so dass er immer heller klingt.

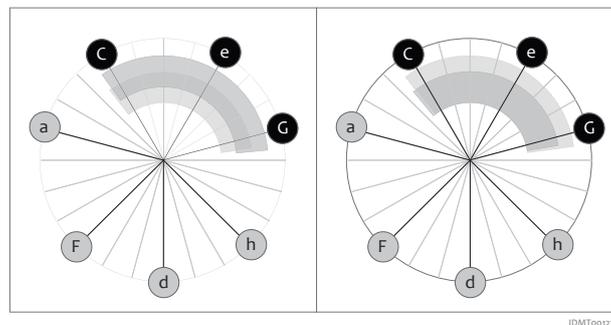


Abbildung 6.7 – a) Eine Veränderung des Startradius blendet einen Akkord durch verschiedene Umkehrungen. (b) Eine Erweiterung des Öffnungsradius fügt einem Akkord zusätzliche Oktavtöne hinzu.

Durch eine Veränderung des *Öffnungsradius* kann die Oktavbreite des selektierten Akkordes verändert werden. Eine Erhöhung des Parameters führt zur Hinzunahme weiterer Oktavtöne und damit zu helleren und voller klingenden Akkorden. Abbildung 6.7b veranschaulicht, wie ein relativ dünn und dunkel klingender C-Dur-Akkord in einen hell und voll klingenden und mehrere Oktaven umfassenden Akkord erweitert wird.

6.5 Die Benutzerschnittstelle

Der für das eigentliche Spiel entscheidende Teil eines Tonraumsynthesizers ist die Benutzerschnittstelle. Diese entscheidet, in welcher Form die Parameter des Tonraumes und der Tonselktionsfunktion dem Spieler präsentiert und von diesem gesteuert werden. In den vergangenen Jahren wurden – bedingt durch die starke Verbreitung elektronischer Spiele, durch die Verfügbarkeit günstiger Sensoren und durch einfach programmierbare Logikschaltkreise – viele Möglichkeiten für neuartige Musikcontroller entwickelt. Das Ziel dieses Kapitels besteht nun darin, am Beispiel von vier solchen Controllern Möglichkeiten zur Steuerung eines Tonraumsynthesizers zu demonstrieren. Bei den verwendeten Controllern handelt es sich um 1.) einen Elo Touchscreen¹, 2.) den JazzMutant Lemur [73], 3.) den 3DConnexion SpaceNavigator [2] und 4.) die Nintendo WiiMote [117].

6.5.1 Der Elo Touchscreen

Der Elo Touchscreen ist ein berührungsempfindlicher VGA-Bildschirm. Neben seiner Funktion als visuelles Ausgabemedium dient der Bildschirm als Zeigegerät. Durch Berühren des Bildschirms an einer Stelle kann der Nutzer den Mauszeiger positionieren und Mausklick-Aktionen ausführen. Der größte Vorteil des berührungsempfindlichen Bildschirms besteht in der Möglichkeit, direkt mit einer Visualisierung des Tonraumes und der Selektionsfläche interagieren zu können. Ein Nachteil handelsüblicher Geräte besteht in der Einschränkung, dass nur ein Berührungspunkt erfasst wird, was wiederum die Anzahl gleichzeitig steuerbarer Parameter stark einschränkt. Folgende Eigenschaften besitzt die Realisierung des HarmonyPads auf dem berührungsempfindlichen Bildschirm:

- Wie in Abbildung 6.2 dargestellt, werden die Tonräume F und TR auf dem Bildschirm visualisiert.
- Durch Berühren eines Tonartenpunktes im F kann der Spieler die Tonart selektieren. Der im Inneren des F dargestellte TR wird daraufhin auf die neu gewählte Tonart umkonfiguriert.
- Die Größe der Selektionsfläche (Öffnungswinkel und Öffnungsradius, Abschnitt 6.4) wird zunächst in einem separaten Dialogfenster so vorkonfiguriert, dass die Selektionsfläche drei Tonigkeiten sowie einen Bereich von zwei Oktaven umfasst.

¹ <http://www.elotouch.de>

- Durch Berühren des TRs kann der Spieler die Auswahlfläche positionieren. Die Auswahlfläche baut sich dabei ausgehend vom berührten Punkt im Uhrzeigersinn auf, so dass Akkorde stets durch Wahl des Grundtones gespielt werden können.
- Durch gleichzeitiges Berühren des Bildschirms und Ziehen oder Schieben des Fingers kann die Selektionsfläche verschoben werden. Die Auswirkung auf den hörbaren Klang ergeben sich entsprechend Abschnitt 6.4.

6.5.2 Der JazzMutant Lemur

Beim JazzMutant Lemur [73] handelt es sich um einen Midi- und OSC² Controller. Die Bedienelemente des Controllers (Fader, Knöpfe, LEDs etc.) werden dabei auf einem integrierten Bildschirm visualisiert und können über eine mitgelieferte Software flexibel angeordnet und konfiguriert werden. Die Bildschirmeinheit des Lemurs erlaubt die Abfrage von bis zu zehn unabhängigen Berührungspunkten. Dadurch kann der Benutzer mit den zuvor konfigurierten Bedienelementen direkt interagieren und auch mehrere Elemente gleichzeitig bedienen. Abbildung 6.8 zeigt die Implementierung des HarmonyPads auf dem Lemur, welche folgende Besonderheiten aufweist:

- Der tonartbezogene Terzkreis *TR* des HarmonyPads wurde mit Hilfe eines sogenannten „Multiballcontrollers“ realisiert, welcher in Abbildung 6.8-8 dargestellt ist. Beim Multiballcontroller handelt es sich um eine rechteckige Fläche, auf der durch Berühren bis zu zehn Bälle positioniert werden können.
- Die X-Koordinate eines Balles wurde auf den Startradius der Selektionsfläche gemappt. Dadurch steigt die Tonhöhe von Einzeltönen und Akkorden wie beim Klavier von links nach rechts an. Die Y-Koordinate des Balles wurde auf den Startwinkel der Selektionsfläche gemappt. Durch Berühren des Multiballcontrollers und Bewegen des Fingers in Y-Richtung kann in benachbarte Akkorde gewechselt werden.
- Der Winkel 0° (Zentrum des Tonikabereiches, Abschnitt 5.3.2) wurde in die Mitte des darstellbaren Y-Wertebereiches gelegt. Positive Winkel (Dominantbereich) befinden sich oberhalb, negative Winkel (Subdominantbereich) unterhalb des Winkels 0° . Für das diatonische Tonset C-Dur ergibt sich dadurch die in Abbildung 6.8 dargestellte Zuordnung von Tonigkeiten zu Y-Positionen.

² OSC = Open Sound Control³ ist ein Netzwerkprotokoll, das hauptsächlich zur Steuerung von Klang erzeugenden und verarbeitenden Geräten verwendet wird.

- Um das Potential des Multiballcontrollers vollständig auszureizen, wurde der Tonraumsynthesizer so erweitert, dass er die Definition *mehrerer* Selektionsflächen (Abbildung 6.8-6/7) entsprechend Abschnitt 6.4 erlaubt. Jede dieser Flächen kann durch einen der Bälle des Multiballcontrollers unabhängig positioniert werden. Dies ermöglicht das Erzeugen von Mischakkorden und stellt so eine große Erweiterung des musikalischen Spielraumes dar.
- Die Spielbarkeit einer kartesischen Anordnung des TRs hat sich als deutlich besser erwiesen, als die einer polaren Anordnung. Ursache könnte sein, dass geradlinige horizontale und vertikale Fingerbewegungen einfacher auszuführen sind als gekrümmte Radius-Winkel-Bewegungen.
- Problematisch bei der kartesischen Anordnung ist der große Sprung beim Übergang vom Subdominant- in den Dominantbereich des TRs. Auch die häufig vorkommende Mischung von Tonigkeiten aus dem Dominant- und Subdominantbereich (siehe Abschnitt 5.3.3, Charakterdissonanzen) erfordert kompliziertere spielerische Gesten.
- Der tonartübergreifende Quintenzirkel *F* zum Wechsel der Tonarten wurde in Form einer Tastenzeile (Abbildung 6.8-3) realisiert. Jede der dreizehn Tasten repräsentiert eine der Tonarten von *Ges-Dur* bis *Fis-Dur* in Quintreihenfolge. Aufgrund dieser fixen Zuordnung werden diese Tasten auch als *absolute Tonartwechseltasten* bezeichnet.
- In der musikalischen Praxis gibt es viele Fälle, in denen der Spieler die Tonart nicht absolut, sondern relativ zu einer aktuellen Tonart wechseln möchte, z.B. um einen bestimmten terzverwandten Akkord zu spielen (Terzverwandtschaft, siehe Abschnitt 5.3.4). Um solche relativen Tonartwechsel zu ermöglichen, wurde die Tastenreihe (Abbildung 6.8-3) in Form der *relativen Tonartwechseltasten* ein zweites Mal ausgeführt (Abbildung 6.8-4). Die mit „0“ bezeichnete Taste repräsentiert die absolut eingestellte Tonart. Die anderen Tasten -6, -5 etc. ermöglichen die Anwahl von Tonarten, die eine entsprechende Anzahl von Quinten davon entfernt sind.
- Um auch das in Abschnitt 5.3.6 beschriebene *Alterieren von Tonigkeiten* zu ermöglichen, wurde das in Abbildung 6.8-5 dargestellte Tastenfeld implementiert. Dieses Tastenfeld kann in aus acht Tasten bestehende Blöcke (Abbildung 6.8-2) eingeteilt werden. Jeder dieser Blöcke ist hierbei einer Tonigkeit (Abbildung 6.8-1) zugeordnet. Durch Drücken einer der Tasten +1, +2 oder +3 – als *Tonerhöhungstasten* bezeichnet – wird die dem Block zugeordnete Tonigkeit um 1, 2 oder 3 Halbtonschritte erhöht.

Die Betätigung einer der Tasten -1, -2 oder -3, welche als *Tonerniedrigungstasten* bezeichnet werden, führt zu einer Erniedrigung der zugeordneten Tonigkeit um 1, 2 oder 3 Halbtonschritte.

- Die Veränderung des Öffnungswinkels und des Öffnungsradius kann durch Hinzunahme weiterer Bedienelemente erfolgen und soll an dieser Stelle nicht weiter beschrieben werden.

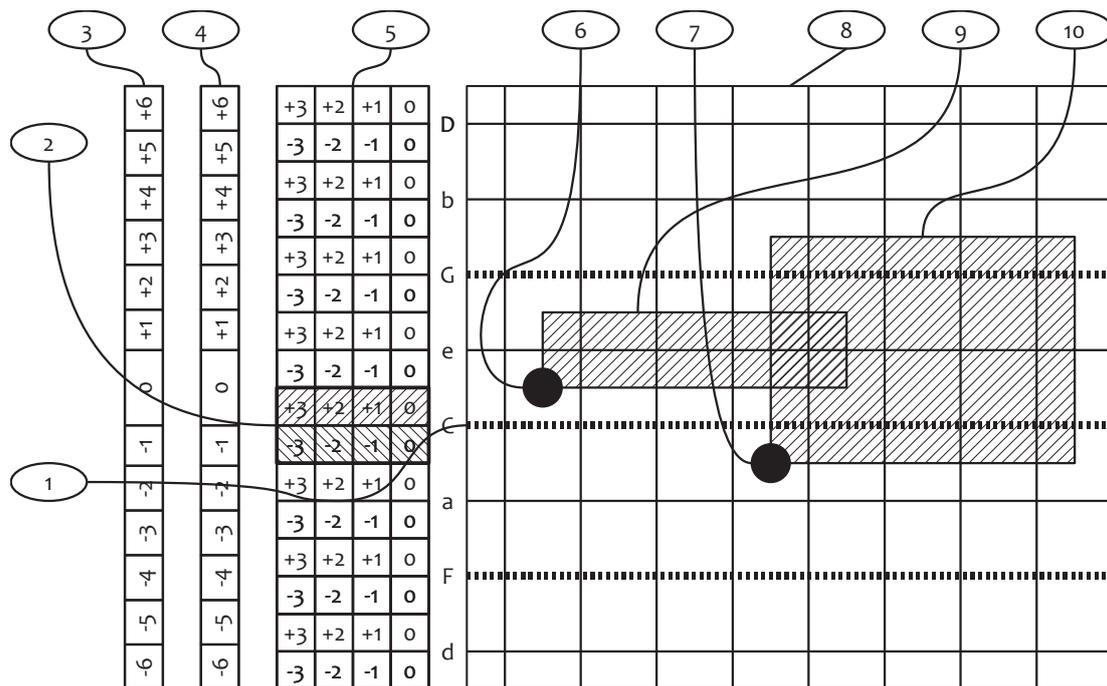


Abbildung 6.8 – Oberflächenlayout des HarmonyPads auf dem Lemur: 1 - Tonigkeitslinie, 2 - einer Tonigkeit zugeordnete Tonerhöhungs- und -erniedrigungstasten, 3 - absolute Tonartwechsellasten, 4 - relative Tonartwechsellasten, 5 - Tonerhöhungstasten, 6,7 - Berührungspunkte auf der Touchoberfläche, 8 - Multitouch-Eingabefläche, 9, 10 - Tonauswahlflächen

6.5.3 3D-SpaceNavigator

Der nachfolgende Abschnitt beschreibt die Verwendung des 3DConnexion SpaceNavigators [2] zur Steuerung des HarmonyPads. Der 3DConnexion Space Navigator ist ein 3D Joystick, der sechs Freiheitsgrade erlaubt: a) Bewegung links/rechts, b) drücken/ziehen, c) nach vorne/hinten neigen, d) rotieren, e) nach rechts/links neigen, f) vorwärts/rückwärts bewegen (Abbildung 8). Da diese Freiheitsgrade verschiedene Operationen im 3D-Raum ermöglichen, soll der SpaceNavigator verwendet werden, um ein Spiel in der dreidimensionalen Erweiterung des tonartbezogenen Terzkreises zu ermöglichen, das ist die in Abschnitt 4.3.1 eingeführte Terzspirale. Die Terzspirale repräsentiert in Form

einer einzelnen Spiralwindung nicht nur sämtliche Eigenschaften des TRs, sondern darüber hinaus werden noch tonartübergreifende Akkordzusammenhänge in geometrisch einfachen Verhältnissen wiedergegeben, was zu einer entsprechend einfachen Spielbarkeit führen kann. Tabelle 6.1 schlägt deshalb ein Mapping vor, das verschiedene Spiralbewegungsoperationen mit Hilfe des SpaceNavigators ermöglicht.

Die Grundidee des Mappings aus Tabelle 6.1 besteht darin, dass das Spiel leitereigener Akkorde dem Navigieren innerhalb einer Spiralwindung entspricht und durch entsprechende 2D-SpaceNavigator-Operationen realisiert werden kann. Das Spiel von Akkorden aus anderen Tonarten entspricht einem Sprung in andere Spiralwindungen. Die Verwendung der 3D-Operationen des SpaceNavigators bietet sich deshalb besonders an. Musikalisch besonders bedeutsame Spiralwindungswechsel sind in Tabelle 6.2 aufgelistet. Im Folgenden sollen nun die Tasks in Tabelle 6.1 erläutert werden: Die Tasks 1-4 in Tabelle 6.1 beschreiben, wie Töne aktiviert werden: Der Startradius und der Startwinkel werden durch Bewegen des SpaceNavigators in Richtung des entsprechenden Punktes definiert. Das abrupte Stoppen dieser Bewegung löst das Spielen dieser Töne aus, wobei die Anschlagstärke der gespielten Töne von der resultierenden negativen Beschleunigung abhängt. Die Tasks 5-9 in Tabelle 6.1 zeigen, wie Tonartwechsel, d.h. Wechsel der selektierten Spiralwindung, realisiert werden können. In abendländischer Musik ist der am häufigsten verwendete Spiralwindungswechsel der Sprung in eine Spiralwindung direkt ober- oder unterhalb der aktuellen Spiralwindung. Entsprechend Abschnitt 4.3.1 transformiert ein solcher Windungswechsel einen gegebenen Mollakkord in den gleichnamigen Durakkord und einen gegebenen Durakkord in den gleichnamigen Mollakkord. Der geometrischen Bewegung entsprechend schlagen wir vor, die Drück- und Ziehfunktion des SpaceNavigators zu verwenden, um diesen Task auszuführen. Dadurch ist es z.B. möglich, einen *e*-Moll-Akkord in einen *E*-Dur-Akkord zu verändern. Tasks 6 und 7 schlagen die Verwendung der Funktionen des SpaceNavigators für andere Tonartwechsel vor. Tonartwechsel, die nicht in Tabelle 6.1 beschrieben sind, können durch Kombination der vorgeschlagenen Bewegungen erzeugt werden⁴. Durch eine Zuweisung der Spiralbewegungsoperationen wie in Tabelle 6.1 beschrieben ist es möglich, alle in Tabelle 6.2 aufgelisteten Tonartwechsel auszuführen. Es ist anzumerken, dass alle in Task 5 bis 7 beschriebenen Tonartwechsel temporäre Tonartwechsel sind, d.h. bei Drücken der Kappe wird die parallele Molltonart sofort ausgewählt. Wird die Kappe danach losgelassen, springt die Auswahl sofort in die vorherige Tonart zurück. Dies erlaubt es, leiterfremde

⁴ Z.B. kann ein Tonartwechsel um einen Ganzton (z.B. von *C* nach *D*) durch eine doppelte Neigung oder eine doppelte Drehung erzeugt werden.

Akkorde sehr schnell „auszuborgen“. Um permanent in eine Tonart zu wechseln, müssen die entsprechenden Tasks (Task 5 bis 7) ausgeführt und danach Taste B gedrückt werden.

	Task	Space Navigator
1	Auswahl des Startwinkels	Bewegen der Kappe in Richtung des gewünschten Winkels
2	Auswahl des Startradius	Bewegen der Kappe zum gewünschten Radius
3	Spielen der selektierten Töne	Stoppen der Bewegung
4	Ändern der Anschlagstärke	Beschleunigung beim Stoppen
5	Sprung in die darüber- bzw. darunterliegende Spiralwindung	Drücken/Ziehen der Kappe
6	Schrauben der Spiralwindung um eine Quinte	Schrauben der Kappe nach links bzw. rechts
7	Schrauben der Spiralwindung um 7 Quinten (± 1 Halbton)	Neigen der Kappe
8	Ausrichtung des Tonraumes auf einen gewünschten Grundton	Physikalisches Rotieren des gesamten SpaceNavigators
9	Permanenter Wechsel in eine andere Tonart	Auswahl der jeweiligen Tonart (Task 5-8) und Drücken der Taste B
10	Definition eines Öffnungsradius	Taste A + Schrauben der Kappe
11	Definition eines Öffnungswinkels	Taste A + Neigen der Kappe

Tabelle 6.1 – Navigation im Tonraum durch Verwendung des 3DConnexion SpaceNavigators

6.5.4 Nintendo Wiimote

Eine der interessantesten Innovationen auf dem Gebiet der elektronischen Spiele war die Einführung der Spielkonsole Nintendo Wii [117] Ende 2006. Das Gerät wird mit einem Controller, der sogenannten Wiimote, ausgeliefert. Diese ist mit Beschleunigungssensoren ADXL330 der Firma Analog Devices [3] ausgestattet, welche die Detektion von Bewegungen sowie die Erkennung der Ausrichtung des Controllers in Ruhelage ermöglichen. Die integrierte Infrarotkamera erlaubt die Ortung von bis zu 4 Infrarotquellen, was eine Bestimmung der räumlichen Position des Controllers ermöglicht. Vier integrierte LEDs, ein



Abbildung 6.9 – Die sechs Freiheitsgrade des 3DConnexion SpaceNavigators [2].

	Tonartwechsel	Spiralwindungswechsel
1	In die gleichnamige Durtonart springen	Sprung in die Spiralwindung direkt über der aktuellen
2	In die gleichnamige Molltonart wechseln	Sprung in die Spiralwindung direkt unter der aktuellen
3	Erhöhung der Tonart um einen Halbton	Spiralwindung um fünf Quinten „herausschrauben“.
4	In eine benachbarte Tonart wechseln	Spiralwindung um eine Quinte „heraus- oder hereinschrauben“

Tabelle 6.2 – Verschiedene Tonart- bzw. Spiralwindungswechsel

Lautsprecher sowie ein Vibrationsmotor geben dem Spieler optisches, akustisches und auch haptisches Feedback. Diese und weitere hier nicht genannte Features machen die Wiimote zu einem sehr attraktiven und innovativen Controller. Die Verfügbarkeit von Open-Source-Treibern, welche z.B. im Zuge des WiiLi Projekts [4] entwickelt wurden, das Linux auf die Wii Konsole portieren will, geben jedem Entwickler die Möglichkeit, die Wiimote in eigene Projekte einzubinden. Beispiele für Musikprojekte sind z.B. die Wii Loop Machine [150] oder das auf der NIME vorgestellte Sensillum [162].

Aufgrund ihrer vielfältigen Eigenschaften ist die Wiimote ein potentieller Controller zur Steuerung des HarmonyPads. Im Rahmen der Forschungsarbeiten wurden deshalb zwei verschiedene Varianten implementiert. In der ersten Variante kann der Spieler durch Drehen der Wiimote um die Y-Achse (Abbildung 6.10, Roll) den Startwinkel der Selektionsfläche des HarmonyPads steuern. Durch Drehen der Wiimote um die X-Achse kann der Startradius kontrolliert werden (Abbildung 6.10, Pitch). In der zweiten Variante verwendet der Spieler die Wiimote als Zeigegerät. Wie mit einem Laserpointer kann er auf eine Visualisierung des HarmonyPads zeigen und somit Startwinkel und Startradius bestimmen. Zum Einstellen des Öffnungswinkels konnte der Spieler mit Hilfe des Steuereckreuzes (Abbildung 6.10, 1) voreingestellte Konfigurationen abrufen. Die Wiedergabe der selektierten Töne erfolgt in beiden Varianten durch Betätigung der Taste A (Abbildung 6.10, 2).

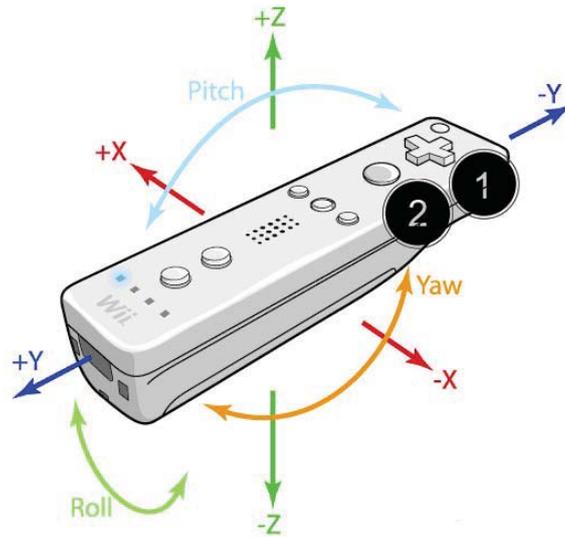


Abbildung 6.10 – Die Wiimote mit eingezeichneten Bewegungsachsen: 1 - Steuerkreuz, 2 - Taste A [116]

6.6 Evaluierung

Um erste Erfahrungen mit dem hier vorgestellten Instrumentenkonzept zu gewinnen, wurden verschiedene informelle Nutzertests, eine Fokusgruppe sowie ein Usabilitytest durchgeführt.

Das Ziel der *Fokusgruppe* bestand darin, das Konzept des HarmonyPads erstmalig einer Gruppe von Testern vorzustellen, sowie das in Tabelle 6.1 vorgeschlagene SpaceNavigator Mapping zu testen. Die Fokusgruppe bestand aus fünf Teilnehmern im Alter zwischen 25 und 29 Jahren, darunter ein klassischer Gitarrist, ein Produzent im Bereich elektronischer Musik, ein Produzent im Bereich Popmusik, ein DJ sowie ein Student mit Erfahrungen im Bereich elektronischer 3D-Eingabegeräte und Usability. Dabei zeigte sich, dass die Freiheitsgrade des SpaceNavigator nicht unabhängig voneinander verwendet werden können. Die Manipulation eines Parameters A führt in vielen Fällen auch zur ungewollten Manipulation eines Parameters B. So kam es vor, dass eine Bewegung der Kappe in X-Richtung gleichzeitig auch eine Bewegung in Z-Richtung und damit einen Tonartwechsel verursachte. Weiterhin führte die gewollte gleichzeitige Steuerung zweier Parameter zu einer hohen Belastung in Armgelenk und Hand und wurde als anstrengend empfunden. Die Ergebnisse der Fokusgruppe zeigten, dass der Test komplexerer Tasks wie Tonartwechsel etc. längerer Übung und weiterer Verbesserungen im Mapping bzw. der Controllerauswahl bedarf. Dies führte zu der Entscheidung, die weiteren Tests auf einfache und sehr grundlegende Tasks zu beschränken.

Das Ziel des an [86] und [40] angelehnten *Usabilitytests* bestand darin, mit Hilfe eines interpersonellen, komparativen Tests die drei auf dem SpaceNavigator, Wiimote und Lemur basierten Varianten des HarmonyPads zu vergleichen, Anforderungen abzuleiten und konkrete Informationen für die Weiterentwicklung zu gewinnen. Folgende Fragen sollten beantwortet werden:

- Welcher der drei verwendeten Controller 3DConnexion SpaceNavigator, Nintendo Wiimote und Jazzmutant Lemur ist am meisten für eine Navigation in Tonräumen geeignet?
- Sind sowohl Musiker als auch Nichtmusiker in der Lage, grundlegende musikalische Tasks im HarmonyPad auszuführen? Ist das Erlernen der Controller einfach und intuitiv?
- Welche Verbesserungsmöglichkeiten ergeben sich für die implementierten Systeme?

6.6.1 Testaufgaben

- Task 1: Spielen von Einzeltönen: Einer Testperson wird eine Visualisierung des Tonraumes gezeigt. In diesem werden einzelne Töne markiert. Die Aufgabe der Testperson besteht darin, die markierten Töne mit Hilfe der gegebenen Controller zu spielen.
- Task 2: Spielen eines Dreiklages: Der Testperson wird eine Visualisierung einer Tonauswahlfläche mit einem Öffnungswinkel eines Dreiklages gezeigt. Aufgabe der Testperson ist es, zunächst den Öffnungswinkel der Tonauswahlfläche einzustellen und anschließend den markierten Akkord zu spielen.
- Task 3: Zeitliche Genauigkeit des Spiels: Es erklingt ein Metronom. Die Aufgabe der Testperson besteht darin, einen Dreiklang im Takt zum gegebenen Metronom zu spielen.
- Task 4: Einstellen des Öffnungswinkels: Der Testperson werden Visualisierungen von Tonauswahlflächen mit verschiedenen Öffnungswinkeln gezeigt. Aufgabe der Testperson ist das Einstellen der gezeigten Öffnungswinkel mit Hilfe der gegebenen Controller.
- Task 5: Zeitlich präzises Spielen einer Akkordfolge: Auf Basis einer vorgegebenen Visualisierung von Akkorden im Tonraum soll die Testperson eine Akkordfolge zum Takt eines Metronoms spielen.

- Task 6: Intuitives Finden von passenden Akkorden zu einem vorgegebenen Klang: Die Testperson hört vorgegebene Töne und soll dazu mit Hilfe der gegebenen Controller jeweils drei passende Akkorde spielen.
- Task 7: Intuitives Finden von passenden Akkorden zu einer vorgegebenen Dreiklangstonfolge: Die Testperson hört einen Dreiklang und soll dazu mit Hilfe der gegebenen Controller zwei ähnliche Akkorde finden.

Die Effektivität der HarmonyPad-Varianten soll durch eine genaue Beobachtung der Testpersonen während der Aufgaben gemessen werden. Es wird protokolliert, inwiefern die gegebenen Aufgaben erfolgreich durchgeführt werden.

Die Effizienz der HarmonyPad-Varianten soll durch die Messung der Lösungsdauer sowie einer Messung der zeitlichen Präzision ermittelt werden. Weiterhin werden sogenannte Critical Incidents gemessen. Das sind besondere Ereignisse während des Tests: Wie oft verliert eine Testperson die Kontrolle? Ein Kontrollverlust wird z.B. durch verbale Äußerungen oder verzweifeltes Verhalten sichtbar.

Die Zufriedenheit der Testpersonen wird über die subjektive Bewertung der Controller anhand mehrerer Likert-Skalen und ein abschließendes vergleichendes Bewerten gemessen. Zudem wurde die Think-Aloud-Methode angewendet: Testpersonen wurden ermutigt, ihre Gedanken während des Tests frei zu äußern. Eine Aufzeichnung erfolgte neben einer schriftlichen Protokollierung per Videokamera.

6.6.2 Verwendetes Parametermapping

Folgendes Mapping wurde in der Usability-Studie verwendet:

Wiimote:

Startwinkel: Zeigen auf den Bildschirm

Startradius: Zeigen auf den Bildschirm

Töne anspielen: Druck auf Taste A/B

Öffnungswinkel: Steuerkreuz für Presets

SpaceNavigator:

Startwinkel: Bewegung in der Ebene

Startwinkel: -

Töne anspielen: Auslenkung der Kappe über einen Schwellwert

Öffnungswinkel: Drehen + B Taste

Lemur:

Startwinkel: Winkel auf einer XY-Eingabefläche

Startradius: Radius auf einer XY-Eingabefläche

Töne anspielen: Berührung der XY-Eingabefläche

Öffnungswinkel: Tasten mit voreingestelltem Öffnungswinkel

6.6.3 Testaufbau

Die Controller SpaceNavigator, Wiimote und Lemur wurden mit einem PC verbunden, auf welchem die Software HarmonyPad installiert ist. Die Visualisierung des Tonraumes erfolgte über einen angeschlossenen Monitor. Die Ausgabe der Toninformation aus dem HarmonyPad erfolgt durch Midisignale, welche über eine Firewire-Soundkarte an einen Laptop mit installierter Klangerzeugung übertragen wurden. Die Firewire-Soundkarte dient gleichzeitig zur Ausgabe der Ton- und Metronomsignale über angeschlossene Lautsprecher. Auf dem Laptop ist neben der Klangerzeugung ein Programm zum Aufzeichnen der Midisignale sowie der Messung der zeitlichen Genauigkeit installiert.

6.6.4 Testdurchführung

Nach einer Begrüßung, Lesen der Teilnehmerinformation, Unterschreiben der Datenschutzeinverständniserklärung, dem Ausfüllen des demografischen Fragebogens bekamen die Testpersonen eine allgemeine Taskbeschreibung zu lesen. Danach wurde die Testreihenfolge der Controller per Lateinischem Quadrat ausgewählt. Danach wurde der Testperson der erste der drei Controller sowie eine Anleitung dazu ausgehändigt. Nach einer kurzen Eingewöhnungszeit führte die Testperson die zuvorbeschriebenen Tasks mit dem gewählten Controller aus und bewertete danach den Controller durch Ausfüllen eines Fragebogens. Dies wurde für die beiden anderen Controller wiederholt. Nach dem letzten Fragebogen sollte noch ein vergleichender Fragebogen ausgefüllt werden.

6.6.5 Ergebnisse

An dem Test nahmen insgesamt 20 Personen teil, davon waren 10 Musiker und 10 Nichtmusiker, 4 weiblich, 16 männlich. Als Musiker werden Personen bezeichnet, die 5 oder mehr Jahre musikalische Erfahrungen aufweisen. Während der Durchführung des Testes musste der Testleiter einige Male helfend eingreifen, jedoch brauchte keiner der Tasks abgebrochen werden. Im Folgenden werden die wichtigsten Ergebnisse der Studie besprochen.

Auswertung der zeitlichen Präzision

Das Rhythmusgefühl der Testpersonen variierte so stark, dass Rückschlüsse auf eine Performance der Controller nicht möglich waren. Teilweise schienen die Testpersonen gar nicht in der Lage zu sein, ihr Spiel am Metronom auszurichten. Eine weitere Auswertung der zeitlichen Genauigkeit soll deshalb hier nicht erfolgen.

Auswertung der Zeitdauer der Tasks

Abbildung 6.11 zeigt Boxplots der Zeiten in Sekunden für Task 1. Es ist zu sehen, dass die Zeiten für alle drei Controller recht ähnlich sind. Ein signifikanter Unterschied zwischen den Controllern ist nicht zu erkennen. Allein in der Gruppe der Musiker wurden mit dem Lemur signifikant bessere Zeiten als mit dem SpaceNavigator erreicht.

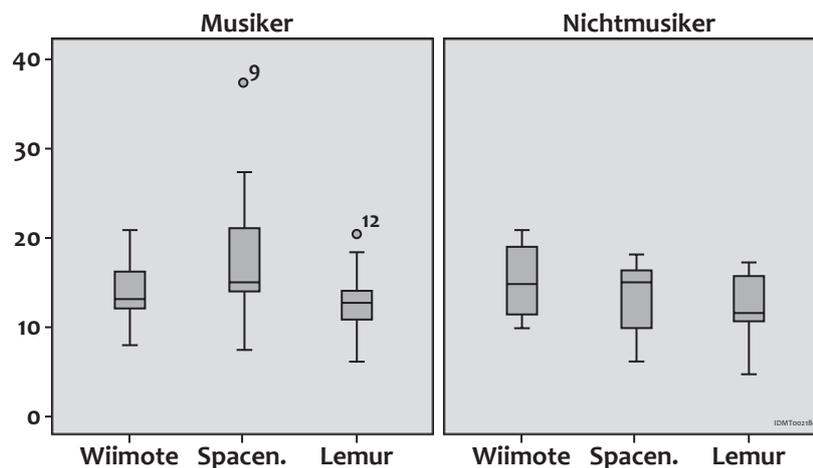


Abbildung 6.11 – Dauer der Durchführung von Task 1 in Sekunden [156]

In Abbildung 6.12 ist die Zeitdauerwertung für Task 4 zu sehen. Hier ergibt sich ein deutlicher Unterschied zwischen den Controllern: In der Gruppe der Nichtmusiker schnitt der Lemur signifikant besser ab als die anderen beiden Controller. Die Gruppe der Musiker war mit dem Lemur schneller als mit dem SpaceNavigator und tendenziell schneller als mit der Wiimote.

Anzahl der Kontrollverluste

Bei sieben Personen kam es mit der Wiimote zu Kontrollverlusten. Eine zusätzliche Einweisung für den Controller war notwendig. Als problematisch erwies sich die Ausrichtung der Wiimote an der Infrarotleiste. Diese deckt nur einen begrenzten Winkelbereich ab, was dazu führt, dass Testpersonen schnell auf Positionen außerhalb des möglichen

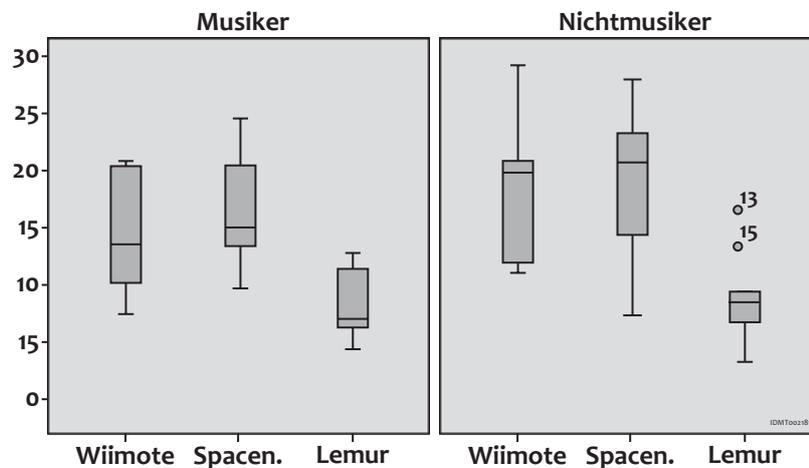


Abbildung 6.12 – Dauer der Durchführung von Task 4 in Sekunden [156]

Winkelbereiches zeigten. Weiterhin kam es mehrfach zu Kontrollverlusten am Lemur. Es stellte sich heraus, dass die Touchoberfläche nicht an allen Stellen korrekt arbeitet. Weiterhin darf der Berührungsdruck nicht zu stark oder zu schwach sein.

Auswertung der Fragebögen

Die Auswertung der Fragebögen ergab eine klare Rangfolge: Platz 1 erreichte der Lemur, gefolgt vom SpaceNavigator auf Platz 2 und der Wiimote auf Platz 3.

Abbildung 6.13 zeigt die Ergebnisse für die Frage, wie gut Töne im Tonraum getroffen werden konnten. In der Gruppe der Musiker bewerteten 8 von 10 Personen den SpaceNavigator mit Höchstnote. Lemur ist hier signifikant besser als SpaceNavigator und Wiimote. In der Gruppe der Nichtmusiker ist die gleiche Tendenz erkennbar, jedoch variieren hier die Ergebnisse stärker.

Eine weitere Frage beschäftigte sich mit dem Einstellen des Öffnungswinkels. Wie in Abbildung 6.14 zu sehen, schnitt hier der SpaceNavigator deutlich schlechter ab, als die anderen beiden Controller. Was die Einfachheit des Erlernens der Controller betrifft, so schnitt der Lemur sowohl bei Musikern als auch bei Nichtmusikern signifikant besser ab. SpaceNavigator und Wiimote wurden in etwa gleich wahrgenommen (Abbildung 6.15).

Auswertung der Tasks 6 und 7

In den Tasks 6 und 7 ging es darum, herauszufinden, inwiefern Testpersonen zu hörbaren Tönen ohne weitere Instruktionen passende Begleitakkorde auf dem HarmonyPad finden konnten. Abbildung 6.16 zeigt die Häufigkeit von Begleitakkorden für den vorgegebenen

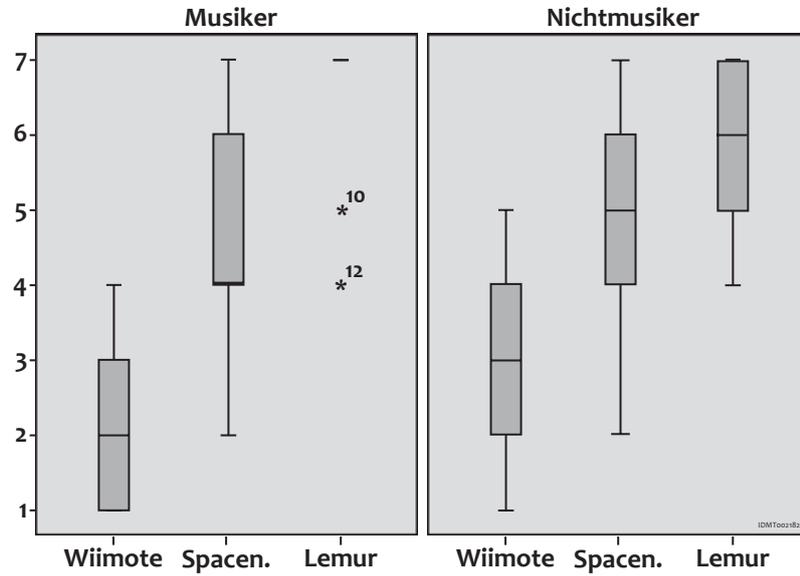


Abbildung 6.13 – Wie gut wurden die Töne im Tonraum getroffen? [156]

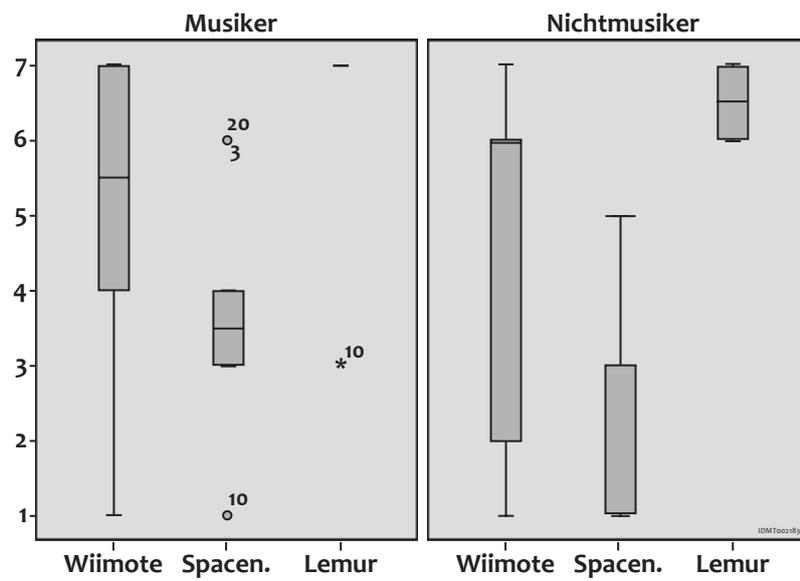


Abbildung 6.14 – Wie leicht ist das Einstellen des Öffnungswinkels? [156]

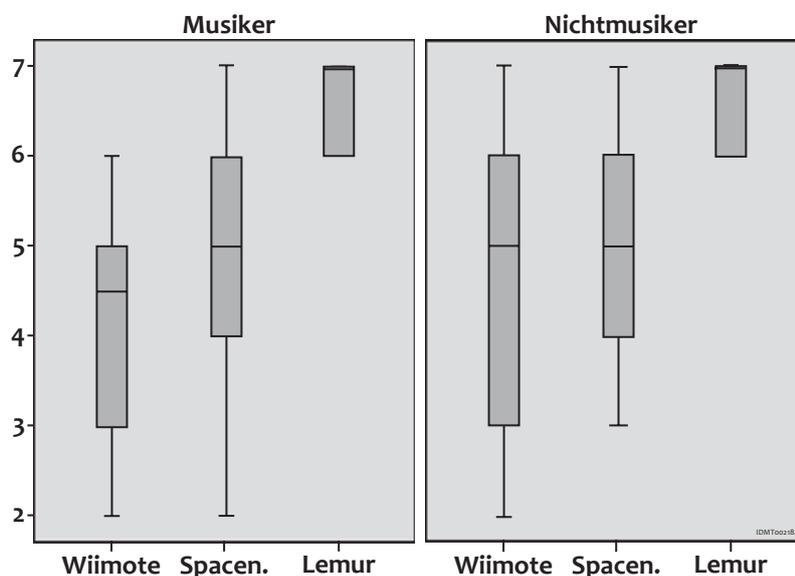


Abbildung 6.15 – Wie leicht ist der Controller zu erlernen? [156]

Ton c. Es ist zu sehen, dass der Ton c am häufigsten mit den Akkorden F und C begleitet wurde. Am seltensten wurde der Ton C mit dem Akkord h-vermindert begleitet. Interessant ist, dass Nichtmusiker als zweithäufigsten Akkord den Akkord e-Moll auswählten. Dieser enthält den vorgegebenen Ton c nicht, ergibt jedoch die interessant klingende Terzschichtung C-e-G-h (auch C-Major Akkord). Generell muss aber gesagt werden, dass bei Nichtmusikern der Unterschied zwischen musiktheoretisch passenden Dreiklängen und weniger passenden Dreiklängen recht schwach ist. Nichtmusiker können zwar recht einfach Akkorde ohne größere Übung spielen, die Auswahl passender Akkorde muss aber durch die Vermittlung von musiktheoretischem Wissen und von Hörtraining begleitet werden.

6.6.6 Diskussion

Die Studie ergab, dass der Jazzmutant Lemur der am besten geeignete Controller ist. Das Mapping zwischen Eingabegeste und Tonraum ist hier am engsten. Eine Verbesserung kann darüberhinaus noch durch eine direkte Visualisierung des Tonraumes auf dem Multitouchscreen erreicht werden – ein Wunsch, der mehrfach geäußert wurde. Haptisches Feedback und Anschlagstärke fehlen auf dem Lemur, was jedoch während des Tests von keiner der Testpersonen vermisst wurde. Der SpaceNavigator ist tendenziell besser als die Nintendo Wiimote. Als Hauptproblem wurde hier das Einstellen des Öffnungswinkels identifiziert. Der Nutzer muss in einen anderen Modus wechseln, was zu lange dauert

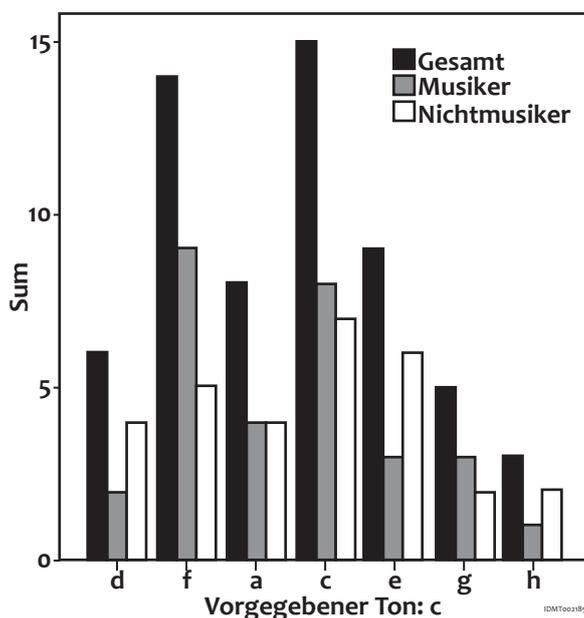


Abbildung 6.16 – Mit Hilfe welcher Klänge begleiteten die Testpersonen den Ton C [156]

und den Spielfluss unterbricht. Auch ist die Auslenkbarkeit der Kappe des Controllers sehr klein, so dass selbst kleine Bewegungen zu größeren Parameter-Veränderungen und damit zu einer mangelnden Kontrolle führen. Die Wiimote ergab sich als der ungeeignetste Controller, trotzdem wurde ihm hoher Spaßfaktor und Intuitivität zugeordnet. Hauptproblem des Controllers ist der enge Winkelbereich des Infrarotsensors. Wird dieser in Zukunft bis hin zu völliger Bewegungsfreiheit erweitert, so sollte sich auch die Wiimote zu einem ernstzunehmenden Controller weiterentwickeln. Insgesamt muss gesagt werden, dass die hier implementierten Mappingkonzepte recht unterschiedlich sind und auch unterschiedliche Schwierigkeitsgrade aufweisen. Hier muss über Lösungen nachgedacht werden, welche die Mappings von der Schwierigkeit her aneinander angleichen. Das gute Abschneiden des Lemurs ermutigt, die Weiterentwicklung des HarmonyPads auf multitouchbasierte Controller zu konzentrieren. Die Schwächen des Lemurs (teilweise fehlerhafte Verarbeitung von Toucheingaben und mangelnde Visualisierungsmöglichkeiten) werden in Zukunft behoben werden. Die Möglichkeiten für multitouchbasierte Musikinstrumente werden sich dadurch noch einmal deutlich erweitern. Detaillierte Beschreibung der Ergebnisse, Fragebogenmaterial etc. wurden ausführlich in [156] dokumentiert.

6.7 Anwendungsmöglichkeiten

Die vorgeschlagene tonraumbasierte musikalische Benutzerschnittstelle ist für verschiedene Zielgruppen interessant. *Kinder* besitzen oft ausgereifte Computerspielerfahrung und können gut mit neuen Hardwarecontrollern umgehen. Ein tonraumbasierter Synthesizer – wie hier präsentiert – kann ihnen helfen, Schritt für Schritt in die Prinzipien abendländischer Musik eingeführt zu werden. Kombiniert mit einer angemessenen Visualisierung können sie schnell musiktheoretische Zusammenhänge erfassen wie den Aufbau von Akkorden, funktionelle Beziehungen zwischen Akkorden (Subdominante, Tonika, Dominante), die Verwandtschaft von bestimmten Tonarten, aber auch psychophysikalische Zusammenhänge wie die Unterscheidung zwischen Tonigkeit, welche dem Winkel zugeordnet ist, und der Tonhöhe, welche dem Radius zugeordnet ist. *Musikstudenten* müssen oft viele musiktheoretische Begriffe lernen. Hierbei ist es vor allem wichtig, den Überblick zu behalten. Die Herausforderung besteht in diesem Zusammenhang in der Überbrückung der Kluft zwischen theoretischem Wissen und praktischer Anwendung. Der Einsatz eines tonraumbasierten Musikinstrumentes könnte Studenten helfen, theoretische Begriffe mit einer räumlichen Vorstellung zu verbinden. Die direkte Interaktion mit einem geometrischen Tonmodell und das sofortige Hörbarmachen der darin verborgenen Tonbeziehungen unterstützen den Lernfortschritt dabei zusätzlich. Aus diesem Grund sollte die vorgeschlagene Benutzerschnittstelle Teil jeder schulischen Musikausbildung sein. *Ältere Menschen* sind oft bereit, ein neues Instrument zu erlernen, jedoch sind klassische Instrumente wie das Klavier oder die Violine zu kompliziert und erfordern die Entwicklung ausgereifter motorischer Fähigkeiten. Ein Musikinstrument, das mit einem sehr einfachen Parametersatz auskommt, könnte ältere Menschen motivieren, eine solche Herausforderung anzunehmen und ein neues Musikinstrument zu erlernen. *Musiker, DJs, Komponisten* können das HarmonyPad als ein innovatives Kreativwerkzeug verwenden, das die Komposition neuartiger, durch alle Tonarten führender Akkordverläufe ermöglicht und damit neue und interessante musikalische Klänge hervorbringt.

Kapitel 7

Grundlagen Musikanalyse

In diesem Kapitel werden die Grundlagen erläutert, die zur Analyse von harmonischen Verläufen in *realem Audio* notwendig sind. Der Begriff *reales Audio* wird in dieser Arbeit für digitale Audiosignale verwendet, wie sie z.B. auf einer Audio-CD zu finden sind.¹ Abbildung 7.1 zeigt den Ablauf der Musikanalyse in dieser Arbeit. Die Analyse lässt sich in drei Bereiche unterteilen:

1. Frontend,
2. Tonraummodell,
3. Analysealgorithmus.

Das *Frontend*² berechnet aus dem zu analysierenden Musikstück eine Folge von sogenannten *Merkmalsvektoren*, die so konstruiert werden, dass sie musikalische Eigenschaften zusammenfassen. Als Merkmalsvektor wird hier der sogenannte *Chromavektor* verwendet, der einen 12-D-Merkmalraum aufspannt. Die in dieser Arbeit angewendete Berechnung des Chromavektors wird in Abschnitt 7.2 beschrieben. In den vergangenen Jahren wurden viele Untersuchungen durchgeführt, um das richtige Berechnungsverfahren für den Chromavektor zu finden, wobei es „das richtige“ nicht gibt, da dieses von der Anwendung³ abhängt. Am Ende dieses Kapitels werden einige Beispiele dafür genannt.

¹ Im Gegensatz zu realem Audio gibt es objektbasierte Audiosignale wie z.B. MIDI-Daten, die nicht das erklingende Audiosignal an sich beschreiben, sondern die zu spielenden Noten und Instrumente.

² Engl. „frontend“ – vordere Begrenzung; das Frontend eines Gesamtsystems ist der Teil, der das System nach vorn abschließt bzw. begrenzt. Mit *vorn* ist hier die Schnittstelle zu den digitalen Audiosignalen gemeint.

³ Verschiedene Anwendungen, wie z.B. Harmonieanalyse oder Rhythmuserkennung, bedürfen einer unterschiedlichen Vorverarbeitung, die jeweils abweichende Inhalte im Chromavektor erzeugen.

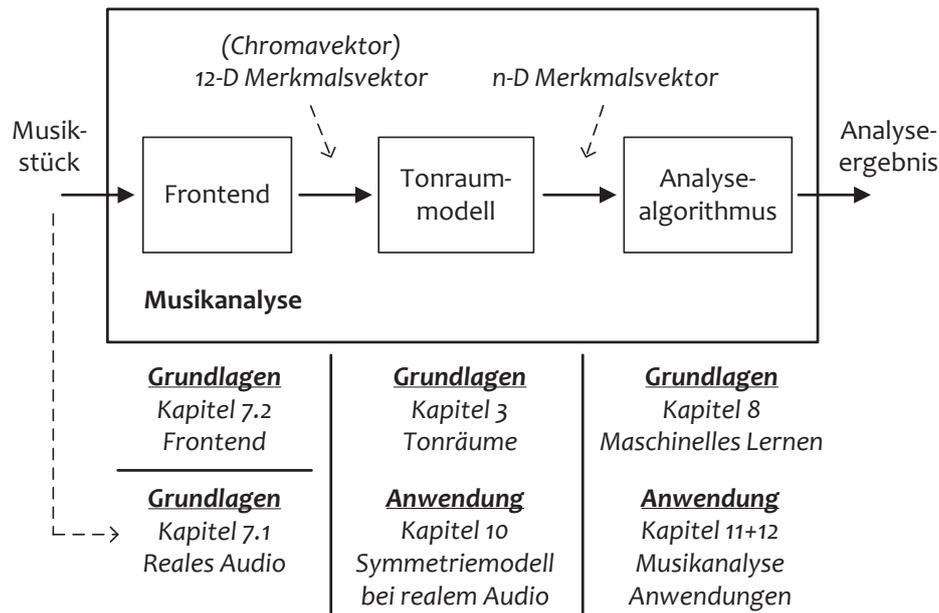


Abbildung 7.1 – Ablauf Musikanalyse

Der Chromavektor stellt ein eigenes Tonraummodell dar und wird als Basis für weitere Tonraummodelle verwendet. Der Block *Tonraummodell* in Abbildung 7.1 berechnet aus den Chromavektoren n -dimensionale Merkmalsvektoren eines anderen Tonraummodells, wie z.B. des Symmetriemodells aus Kapitel 4. In der Gesamtheit aller Merkmale beinhalten die Merkmalsvektoren eines neuen Tonraummodells keine neuen Informationen. Allerdings unterscheiden sich die neuen Einzelmerkmale von denen des Chromavektors und können besser geeignet sein, musikalische Parameter, wie z.B. Harmonie, zu repräsentieren. Die Ausgabeparameter (Merkmale) dieses Modells für reales Audio werden in Kapitel 10 diskutiert.

Der *Analysealgorithmus* ermittelt in Verbindung mit den Merkmalsvektoren des ausgewählten Tonraummodells (z.B. Chromavektor oder Merkmalsvektoren des Symmetriemodells) das Ergebnis der Analyse. Die Grundlagen für die hier verwendeten Analysealgorithmen werden in Kapitel 8 erläutert. Die Anwendungen der in dieser Arbeit vorgestellten Analysealgorithmen für Tonart- und Akkorderkennung sind in den Kapiteln 11 und 12 zu finden. Der nächste Abschnitt 7.1 gibt zuerst einen kurzen Einblick über die Eigenschaften von realem Audio, bevor das Frontend im anschließenden Abschnitt 7.2 erläutert wird.

7.1 Reales Audio – Frequenzanalyse

Dieser Abschnitt beschreibt die Eigenschaften von realem Audio beim Erklngen eines einzelnen Klaviertons. Da Musik aus Tönen (Frequenzgemischen) besteht, basiert die Analyse von Musikstücken auf Algorithmen, die Frequenzabbildungen als Merkmalsvektoren benutzen. Das Diagramm (a) in Abbildung 7.2 zeigt den einzelnen Klavierton als Zeitsignal und zwei Spektren davon zu jeweils verschiedenen Zeitpunkten. Die Einzelereignisse der Musik stehen zwar in einem Frequenzverhältnis, aber die Musik als Ganzes wird in der zeitlichen Abfolge von Einzelereignissen wahrgenommen. Deshalb muss das Musikstück zu verschiedenen Zeitpunkten auf Einzelereignisse (Frequenzanalyse) untersucht werden. Dafür wird zu jedem Zeitpunkt ein Zeitfenster mit definierter Länge aus dem Musikstück ausgeschnitten.

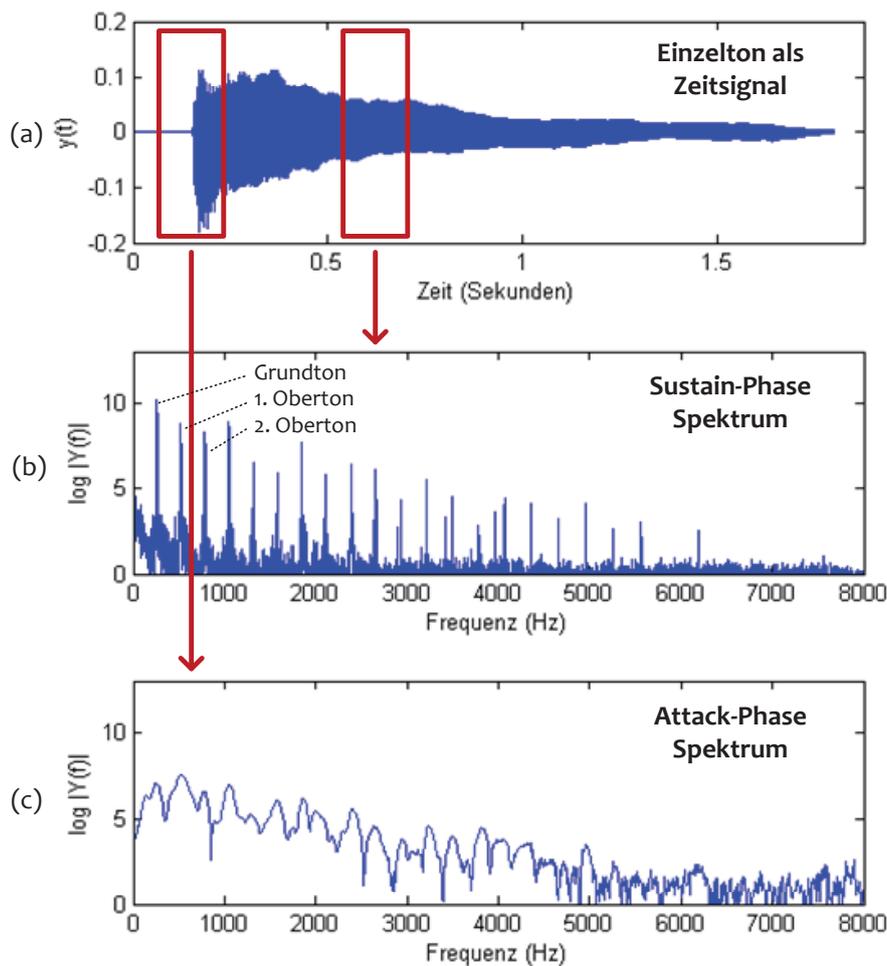


Abbildung 7.2 – Einzelton und dessen Spektren zu verschiedenen Zeitpunkten

Bei der Analyse von Einzeltönen⁴, die für die vollständige Transkription eines Stückes benötigt werden, wird das Musikstück alle 5 – 10 *ms* untersucht. Da in dieser Arbeit Akkorde und Tonarten erkannt werden sollen, die sich in größeren Zeitbereichen ändern, genügt ein Analysewiederholintervall im Bereich von 100 *ms*. Damit einher gehen unterschiedliche Zeitfensterlängen. Im Diagramm (b) in Abbildung 7.2 ist das Spektrum des Tones während der sogenannten „Sustain-Phase“⁵ zu sehen. Die tiefste ausgeprägte Frequenzlinie ist der Grundton und entspricht der wahrgenommenen Tonhöhe. Die weiteren ausgeprägten Frequenzlinien sind die sogenannten „Obertöne“⁶, die den Klang des Klaviertones formen. Jeder natürlich erzeugte Ton besitzt ganzzahlige Vielfache seines Grundtones. Der erste Oberton entspricht der Oktave, der zweite der Quinte, der dritte der Oktave und der vierte der Terz im Verhältnis zum Grundton. Die Obertonreihe kann beliebig fortgeführt werden und es kommen weitere Intervalle hinzu (vgl. [43] S. 71). Jedoch hören wir nicht mehrere Töne beim Erklängen eines einzelnen Klaviertons, da die menschliche Musikwahrnehmung alle Obertöne dem Grundton zuordnet und somit einen einzigen Ton mit einem bestimmten Klang entstehen lässt (vgl. [32] S. 97 ff.). Diese Leistungsfähigkeit konnte für ein beliebiges Audiosignal bisher nicht vollständig algorithmisch beschrieben werden und so erkennen rechnerbasierte Musikanalyseansätze neben dem Grundton zum Teil weitere Töne, nämlich die Obertöne. Allerdings ist der Grundton in der Regel der lauteste Anteil und die Obertöne nehmen in ihrer Intensität ab.

Im Diagramm (c) Abbildung 7.2 ist das Spektrum des Klaviertones während der sogenannten „Attack-Phase“ zu sehen. Da der Klavierton zu diesem Zeitpunkt gerade angeschlagen wird, ist der zeitliche Verlauf der Hüllkurve des Signals impulshaft. Die Energie des gerade angeschlagenen Tones verteilt sich so über das Spektrum, dass die einzelnen Spektralanteile des Tones nicht deutlich zu erkennen sind. Die harmonische Analyse wird an dieser Stelle durch das „rauschhafte“ Spektrum gestört bzw. erschwert. Ähnliche Spektren werden durch Rhythmus-Instrumente – wie z.B. ein Schlagzeug – erzeugt. Da in dieser Arbeit die Zeitfensterlänge wesentlich größer als 100 *ms* ist, tritt der beschriebene Effekt abgeschwächt auf.

⁴ Engl.: „on set detection“ und „off set detection“ – das Erkennen des Beginns und des Endes einer gespielten Note (vgl. [11], [54], [39]).

⁵ Das Erklängen eines Tones setzt sich aus vier Phasen zusammen, engl.: Attack – Decay – Sustain – Release; Anschlagen – Abfallen – Halten – Loslassen. Vgl. [135] ADSR-Modell.

⁶ Obertöne sind alle ganzzahligen positiven Vielfachen des Grundtones ohne den Grundton selbst.

7.2 Frontend

Die Grundlage für die eigentliche Musikanalyse sind die Merkmalsvektoren, die eine bestimmte Anzahl an Einzelmerkmalen (Dimensionen) des Frequenzspektrums zusammenfassen. Der Chromavektor beinhaltet die Amplituden aller 12 Tonigkeiten für das gerade zu analysierende Zeitfenster. In dieser Arbeit wird er in zwei Funktionen verwendet: erstens zur Nachbildung und Verifizierung der Algorithmen, die als Vergleich für die Evaluierung ausgewählt werden, und zweitens als Eingangsvektor für die anderen verwendeten Tonraummodelle, wie z.B. das Symmetriemodell. Abbildung 7.3 zeigt das Blockschaltbild des Frontends zur Erzeugung der Chromavektoren. Als Eingabedaten können *reales Audio* (PCM-Daten⁷ von einer Audio-CD) oder *symbolisches Audio* (MIDI-Daten⁸) verwendet werden. Die einzelnen Funktionsblöcke *Auswahl des Zeitfensters*, *Constant-Q-Transformation*, *Tuning und Mapping* werden in den folgenden Abschnitten genauer erläutert. Der nächste Abschnitt legt zuerst die technischen Parameter des zu analysierenden PCM-Signals fest.

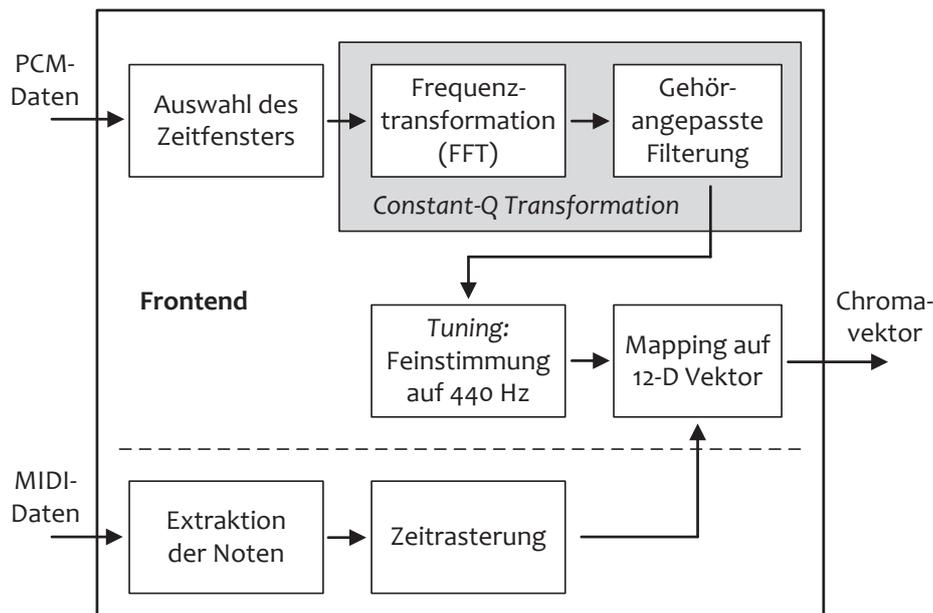


Abbildung 7.3 – Frontend – Berechnen der Chromavektoren

⁷ Engl.: “pulse code modulation” – impuls-kodierte Modulation. Vgl. [75] S. 191.

⁸ „Musical Instrument Digital Interface“. MIDI ist ein Standard zum Ansteuern von digitalen Musikinstrumenten. Vgl. [7].

7.2.1 Eigenschaften PCM-Signal

Das Musiksignal einer Audio-CD ist ein digital gespeichertes PCM-Stereosignal. Die Abtastrate beträgt $44,1\text{ kHz}$ und die Auflösung der einzelnen Abtastwerte 16 Bit . Damit lässt sich ein Audiosignal übertragen, das den gesamten hörbaren Frequenzbereich des Menschen beinhaltet (vgl. [43] S.17 ff.) und einen theoretischen Signal/Rausch-Abstand von $98,1\text{ dB}$ bei einem voll ausgesteuerten Sinussignal hat (vgl. [17] S. 360). Die Stereoanteile enthalten keine Informationen, welche für die harmonische Musikanalyse in dieser Arbeit notwendig sind.⁹ Daher wird das Musiksignal in ein Monosignal umgerechnet. Der zu untersuchende Frequenzbereich wird auf 55 Hz bis 3520 Hz festgelegt und umfasst damit genau sechs Oktaven (72 Halbtöne). Die Abtastrate kann dadurch auf $8,0\text{ kHz}$ reduziert werden. Diese Parameter finden auch in der Literatur häufig Verwendung (vgl. [168], [37], [36]).

7.2.2 Auswahl des Zeitfensters

Die Auswahl des Zeitfensters gibt vor, zu welchen Zeitpunkten (Schubweite) das Musiksignal analysiert werden soll und wie lang ein Zeitfenster (Fensterlänge) ist. Abbildung 7.4 zeigt die Parameter des zu analysierenden Zeitfensters beispielhaft an einem Musiksignalausschnitt.

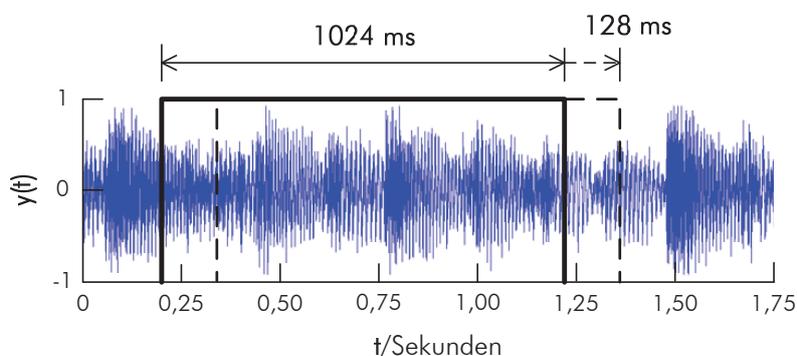


Abbildung 7.4 – Parameter des zu analysierenden Zeitfensters

Die Schubweite darf nicht länger sein als die kürzesten zu erkennenden Ereignisse, welche in dieser Arbeit die Akkordwechsel sind und im Bereich von 300 ms liegen. Für eine gute Zeitauflösung der zu analysierenden Akkorde und Tonarten wird eine Schubweite

⁹ Es ist durchaus denkbar, dass über Panning (engl. für Panoramasteller) ein Rhythmusinstrument, welches für die Akkord- und Tonarterkennung Rauschen darstellt, teilweise herausgefiltert werden kann. Dies ist aber nicht Gegenstand der vorliegenden Arbeit.

von 128 ms festgelegt. Die Fensterlänge beträgt 1024 ms und wird durch die feinste zu analysierende Frequenzauflösung vorgegeben, die einen Drittel Halbton bei 55 Hz groß ist. Im nächsten Abschnitt „Constant-Q-Transformation“ wird dies genauer betrachtet.

7.2.3 Constant-Q-Transformation

Die Constant-Q-Transformation (vgl. [15] S. 425 ff.) erzeugt ein den Tonhöhen und der Tonhöhenwahrnehmung angepasstes Frequenzspektrum. Zur Veranschaulichung wird die Constant-Q-Transformation in zwei Schritten durchgeführt und in Abbildung 7.5 dargestellt.

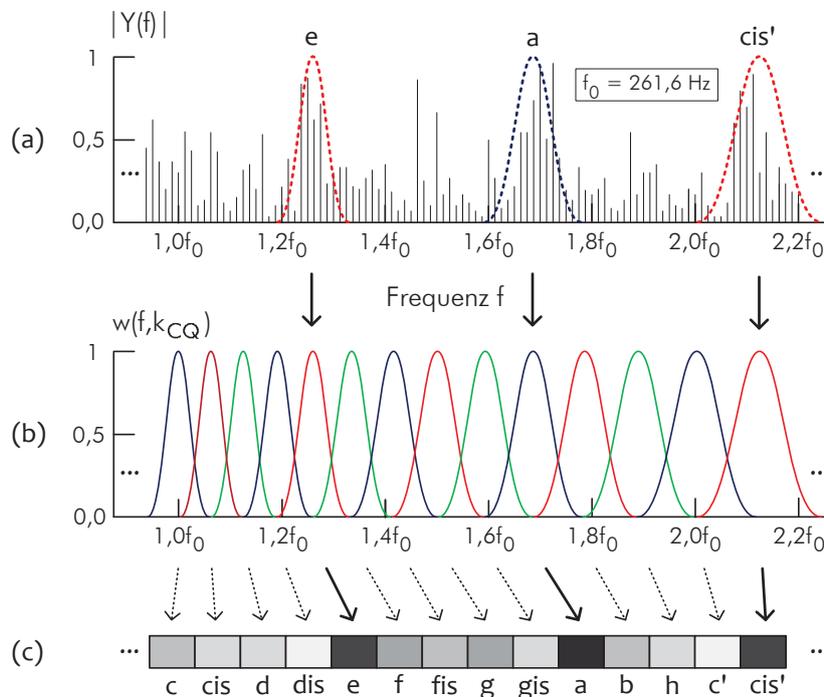


Abbildung 7.5 – Constant-Q-Transformation

Zuerst erfolgt die Frequenztransformation mittels einer FFT.¹⁰ Im Diagramm (a) ist das Ergebnis der FFT zu sehen und soll dem Spektrum des Zeitfensters aus Abbildung 7.4 entsprechen. Anschließend wird dieses Spektrum mit einer Filterbank, deren Filter eine konstante Güte haben (Constant-Q¹¹), gefiltert. Das Diagramm (b) zeigt die Filter mit

¹⁰ Fast Fourier Transformation. Die FFT ist ein effizienter Algorithmus zum Berechnen einer diskreten Fouriertransformation. Vgl. [23].

¹¹ Constant-Q steht auch für „konstanter Quotient“. Damit ist das Verhältnis aus der Tonhöhe und seiner Bandbreite gemeint. Mit steigender Tonhöhe nimmt die Bandbreite des belegten Frequenzbereiches zu; z.B. bei Verdopplung der Tonhöhe verdoppelt sich auch die Bandbreite.

konstanter Güte. Das Ergebnis der Filterung ist ein Tonvektor aus 72 Tönen¹² mit unterschiedlichen Amplituden. Ein Ausschnitt des Tonvektors ist im Diagramm (c) zu sehen. Die Amplituden der Töne *e*, *a* und *cis'* sind am stärksten. Das musikalische Ereignis in dem analysierten Zeitfenster ist mit hoher Wahrscheinlichkeit der Akkord *A-Dur*. In den restlichen nicht dargestellten Tönen könnte allerdings noch ein weiterer lauter Ton, z.B. *Fis*, enthalten sein. Dann wäre der Akkord ein *fis-Moll-Septakkord* und nicht der Akkord *A-Dur*. Die Constant-Q-Transformation ist wie folgt definiert:

$$X[k_{cq}] = \sum_{n=0}^{N[k_{cq}]-1} w[n, k_{cq}] x[n] e^{-j2\pi f_{cq} n} \text{ mit } k_{cq} = 0 \dots (6 \cdot \alpha - 1) \quad (7.1)$$

Der Tonvektor ist $X[k_{cq}]$. Der Faktor α gibt die Anzahl der Töne pro Oktave an und beträgt für das oben gezeigte Beispiel 12. Der Tuning-Block aus Abbildung 7.3 benötigt allerdings 36 Unterteilungen pro Oktave, um die Feinstimmung des Musikstücks ermitteln zu können. Daher wird α auf 36 festgesetzt. $w[n, k_{cq}]$ beschreibt die einzelnen Constant-Q-Filter und $x[n]$ ist das zu analysierende Zeitfenster mit $N[0]$ Abtastwerten. f_{cq} entspricht den Mittenfrequenzen der Filter:

$$f_{cq} = \left(2^{\frac{1}{\alpha}}\right)^{k_{cq}} f_{\min}, \text{ mit } f_{\min} = 55\text{Hz} \quad (7.2)$$

Mit steigender Mittenfrequenz f_{cq} werden weniger Abtastwerte $N[k_{cq}]$ zur Berechnung des gefilterten Signals verwendet. $N[k_{cq}]$ ist invers zur Filterbandbreite. Die meisten Abtastwerte werden damit bei der niedrigsten Frequenz f_{\min} benötigt. Durch die niedrigste Frequenz f_{\min} und die Anzahl der Töne pro Oktave α wird die Zeitfensterlänge t_F vorgegeben:

$$t_F \geq \frac{1}{\left(2^{\frac{1}{\alpha}} - 1\right) f_{\min}} \quad (7.3)$$

$$\Rightarrow t_F \geq \frac{1}{\left(2^{\frac{1}{36}} - 1\right) 55\text{Hz}} \quad (7.4)$$

$$\Leftrightarrow t_F \geq \frac{1}{1,069\text{Hz}} \quad (7.5)$$

$$\Leftrightarrow t_F \geq 0,935\text{s} \quad (7.6)$$

Da zur schnellen Berechnung der Constant-Q-Transformation der Algorithmus aus [15] verwendet wird, muss die Anzahl der Abtastwerte eines zu analysierenden Zeitfensters

¹² 6 Oktaven · 12 Töne/Oktave = 72 Töne.

einer Potenz von 2 entsprechen. Es wird die Anzahl 2^{13} gewählt, was bei einer Abtastrate von 8 kHz einer Zeit von $t_F = 1024\text{ ms}$ entspricht. Diese Zeit erfüllt die Bedingung aus Gleichung 7.6. Als Filterfunktionen $w[n, k_{cq}]$ kommen Hamming-Fenster zur Anwendung. Diese werden wie folgt definiert:

$$w[n, k_{cq}] = \beta - (1 - \beta) \cos \frac{2\pi n}{N[k_{cq}]}, \text{ mit } \beta = \frac{25}{46} \quad (7.7)$$

Durch die feinere Auflösung von 36 Tönen pro Oktave ergeben sich nun 216 Töne für den Tonvektor, der das Ergebnis der Constant-Q-Transformation darstellt.

7.2.4 Tuning

Normalerweise sind Instrumente so gestimmt, dass der Ton a' eine Grundfrequenz von 440 Hz hat (Kammerton a). Allerdings ist dies nicht immer der Fall. Die Stimmung eines Orchesters oder einer Musikgruppe kann von der 440 Hz -Stimmung innerhalb eines Halbtones beliebig abweichen. Digitalisierte Tonbandaufnahmen oder Plattenaufnahmen, z.B. Beatles-Lieder, weichen in der Regel ebenfalls von der 440 Hz -Stimmung ab.

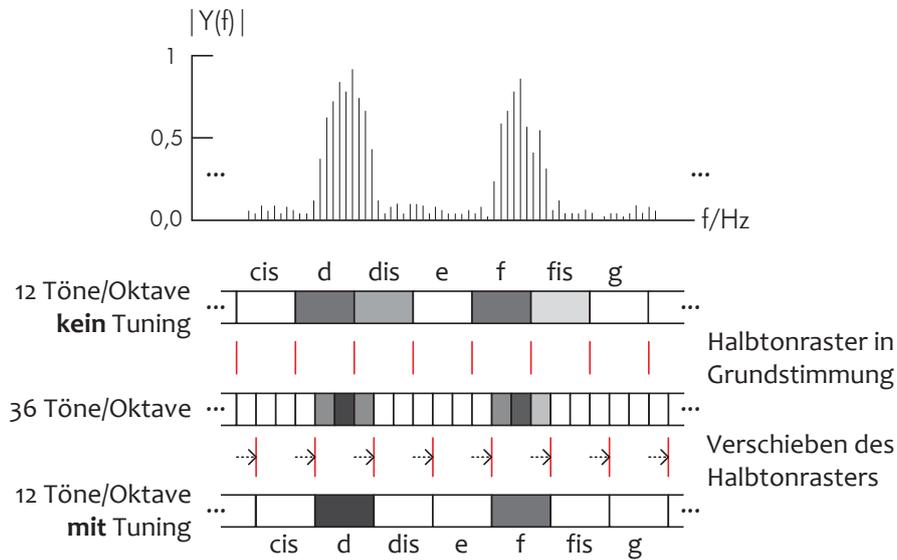


Abbildung 7.6 – Tuning mit 36 Tönen pro Oktave

Abbildung 7.6 zeigt als Beispiel ein Frequenzspektrum, bei dem die Stimmung um einen sechstel Ton von der 440 Hz -Stimmung nach oben abweicht. Dies ist am Ausschnitt des nicht gestimmten Tonvektors (12 Töne/Oktave, kein Tuning) zu erkennen. Hier befindet sich zwar der größere Anteil an Energie in den beiden erklingenden Tönen, allerdings auch ein Anteil in den um einen Halbton versetzten Tönen dis und fis . Dies beeinträchtigt

die nachfolgenden Bearbeitungsstufen und die Leistungsfähigkeit des Gesamtsystems. Wie in Abschnitt 7.2.3 erwähnt, wird für die Constant-Q-Transformation eine Auflösung von 36 Tönen pro Oktave gewählt. Dadurch ist jeder Halbton dreifach unterteilt. Zur Feinstimmung auf 440 Hz werden die Bereiche im feiner aufgelösten Tonvektor gesucht, welche die größte Gesamtenergie beinhalten. Das Halbtonraster wird derart neu berechnet, dass die Bereiche der größten Energie in der Mitte des neuen Rasters liegen. Das Ergebnis ist ein gestimmter Tonvektor mit insgesamt 72 Tönen.

7.2.5 Mapping

Nach dem Tuning beinhaltet der Tonvektor 72 Töne und damit 72 Dimensionen. Dieser könnte als Merkmalsvektor verwendet werden. Da aber die Oktavlage eines Tones für die Harmonieempfindung keine Information trägt¹³, ist es möglich, alle Töne mit der gleichen Tonigkeit zusammenzufassen. Dies wird mit dem Begriff *Mapping* bezeichnet. Das Ergebnis ist der 12-dimensionale Chromavektor. Abbildung 7.7 zeigt die Generierung des Chromavektors beispielhaft an der Tonigkeit a. Der Chromavektor wird vom Frontend ausgegeben und ist die Basis für die Musikanalysealgorithmen.

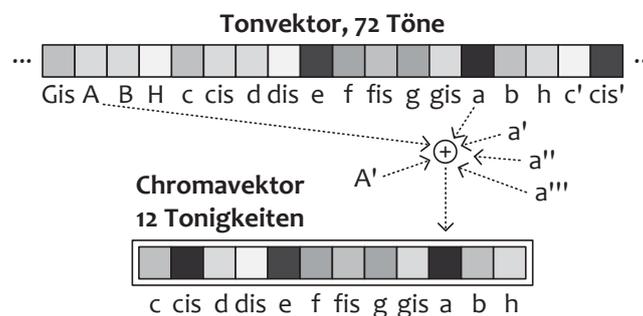


Abbildung 7.7 – Mapping des Tonvektors auf den Chromavektor

7.2.6 MIDI-Datenverarbeitung

Das MIDI-Protokoll (vgl. [7]) kann als Partitur interpretiert werden, das in der Lage ist, Noten für Instrumente jeder Art (auch Schlagzeug) zu beschreiben. Damit enthalten MIDI-Daten bereits die Grundtöne, die vom Frontend in realem Audio gesucht werden. MIDI-Stücke sind daher für das Trainieren und Verifizieren von Analysealgorithmen

¹³ Der tiefste Ton (Basston) hat eine Sonderstellung, da er der Bezugston der Harmonie bzw. des Akkordes ist. Mit ihm kann auch ein mehrdeutiger Akkord, z.B. ein a-Moll-Septakkord, als C-Dur-Sextakkord umgedeutet werden, wenn der Basston wie in diesem Beispiel ein c ist.

geeignet. In dieser Arbeit werden dafür diejenigen Instrumente ausgewählt, die harmonische Informationen tragen, wie z.B. Klavier, Gitarre oder Bass. Die MIDI-Noten werden in 128 *ms*-Zeitbereiche gerastert und in Chromavektoren umgerechnet, was dem Mapping aus Abschnitt 7.2.5 entspricht.

7.2.7 Zusammenfassung der Parameter des Frontends

Das Frontend lässt sich mit folgenden Parametern zusammenfassend beschreiben:

PCM-Signal

Signalart: Mono

Abtastrate: 8 *kHz*

Abtastwertauflösung: 16 *Bit*

Niedrigste Frequenz f_{\min} : 55 *Hz*

Höchste Frequenz f_{\max} : 3250 *Hz*

Zeitfenster

Länge: 8192 Abtastwerte (1024 *ms*)

Schubweite: 1024 Abtastwerte (128 *ms*)

Constant-Q-Transformation

Auflösung: 36 Töne pro Oktave

Filterart: Hamming

Tuning

Art: Maximumbestimmung im drittel Halbtonraster

Mapping und Ausgabe

Mapping: Zusammenfassen der gleichen Tonigkeiten

Ausgabe: Chromavektor, Chromagramm¹⁴

Damit ist das Frontend vollständig beschrieben. Die dargestellten Grundlagen zur Berechnung der Chromavektoren sind Stand der Technik. Neben der Verwendung der Constant-Q-Transformation gibt es weiterhin die Möglichkeit, direkt im FFT-Spektrum eine bestimmte Anzahl an Maxima zu suchen, was mit „Peak-Picking“ bezeichnet wird (vgl. [60]). Mit Hilfe der gefundenen Maxima werden zuerst das Tuning ermittelt und

¹⁴ Eine Matrix aus zeitlich aufeinanderfolgenden Chromavektoren wird als *Chromagramm* bezeichnet.

anschließend die Chromavektoren berechnet. Des Weiteren schlägt Emilia Gómez anstelle von Chromavektoren die Verwendung von „Harmonic Pitch Class Profiles“ (HPCP) vor (vgl. [52]) und Kyogu Lee führt sogenannte „Enhanced Pitch Class Profiles“ (EPCP) ein (vgl. [93]). Beide Ansätze versuchen zur Grundtonbestimmung Obertöne aus dem Spektrum zum Grundton hinzuzuaddieren. Es konnte bisher nicht allgemein gezeigt werden, dass diese beiden Ansätze anstelle von Chromavektoren mit der Verwendung von Musikanalysealgorithmen, z.B. maschinelles Lernen, bessere Ergebnisse erzielen. Die vorliegende Arbeit nutzt zur Evaluierung ausschließlich Chromavektoren, die – so wie hier gezeigt – berechnet werden. Es folgt das Kapitel „Maschinelles Lernen“, das die Grundlagen für die hier verwendeten Musikanalysealgorithmen darstellt.

Kapitel 8

Grundlagen Maschinelles Lernen

In diesem Kapitel werden die notwendigen Grundlagen erläutert, die für das Verstehen und Anwenden von Erkennungsalgorithmen, die auf Basis von Maschinenlernverfahren funktionieren, notwendig sind. Zuerst wird das maschinelle Lernen allgemein erklärt. Danach folgen ein Überblick über die in dieser Arbeit verwendeten Verfahren und anschließend die Einführung eines für das *Music Information Retrieval* im Bereich Akkorderkennung neuen Verfahrens, die Kombination aus *Hidden Markov Models* mit einem *k Nearest Neighbors*-Schätzer.

8.1 Maschinelles Lernen allgemein

Maschinelles Lernen bedeutet, einen Computer so zu programmieren, dass ein zu bestimmendes Leistungskriterium mit Hilfe von Beispieldaten aus der Vergangenheit optimiert wird. Dazu muss ein Modell ausgewählt werden, dessen Parameter beim Lernen (Training) so angepasst werden, dass zukünftige Vorhersagen anhand neuer Daten optimal getroffen werden können (vgl. [8], S.3).

Es gibt viele Teilbereiche, in denen das maschinelle Lernen angewendet wird. Ein prominenter Anwendungsbereich ist die Spracherkennung (vgl. [129]), in die viel Forschungsaufwand gesteckt worden ist und bei der sich *Hidden Markov Models* als geeignetes maschinelles Lernverfahren herausgestellt haben. Die Mustererkennung, zu der die Handschriftenerkennung oder das Finden eines Barcodes in einem Bild gehören, zählt ebenfalls dazu. Ein weitere Beispielanwendung ist das Finden von Assoziationsregeln beim Einkaufen, die ausdrücken können, ob ein Kunde bestimmte Produkte stets zusammen kauft. Dies sind Anwendungen bei dem in erster Linie das sogenannte *überwachte Lernen* zum Einsatz kommt. Hier liegt beim Training eine Übungsmenge mit geordneten Paaren vor, d.h. der lernende Algorithmus kennt sowohl die zu untersuchenden Eingabedaten als

auch deren Zuordnung zu einem Ergebnis, einer sogenannten *Klasse*. Daher wird der Vorgang, neue Eingabedaten einer Klasse zuzuordnen, auch *Klassifizierung* genannt. Das Ergebnis einer abgeschlossenen Klassifizierung heißt *Klassifikation* (vgl. [44], S.174). Im Gegensatz zum überwachten Lernen enthält beim sogenannten *unüberwachten Lernen* die Übungsmenge nur die zu untersuchenden Eingabedaten selbst und keine Zuordnung zu bestimmten Klassen. Das Trainingsziel besteht darin, bestimmte Regelmäßigkeiten in den Eingabedaten aufzuspüren. Ein Beispiel hierfür ist die sogenannte *Clusteranalyse*. In dieser Arbeit kommt ausschließlich das überwachte Lernen zum Einsatz und wird im folgenden Abschnitt näher erläutert.

8.1.1 Überwachtes Lernen

Das überwachte Lernen benötigt geordnete Paare aus Eingabedaten und dazugehörigen Klassen. Für die in dieser Arbeit angewendeten Lernverfahren sind die Eingabedaten die Merkmalsvektoren der ausgewählten Tonraummodelle. Sie werden als Vektor \mathbf{x} definiert und bestehen aus M Attributen in der Form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_M \end{bmatrix} . \quad (8.1)$$

Es gibt K unterschiedliche Klassen, geschrieben als $C_i, i = 1, \dots, K$. Die verwendeten Klassen sind hier die Tonarten und Akkorde. Diese werden durch den Klassenvektor \mathbf{r} mit der Dimension K repräsentiert und es gilt

$$r_i = \begin{cases} 1 & \text{falls } \mathbf{x} \in C_i, \\ 0 & \text{falls } \mathbf{x} \in C_j, j \neq i . \end{cases} \quad (8.2)$$

Der Klassenvektor hat damit an der i -ten Stelle eine 1 und an allen anderen Stellen eine 0, wenn er die Klasse C_i darstellen soll. Zum Trainieren des ausgewählten Modells wird eine Übungsmenge \mathbf{X} mit N geordneten Paaren (\mathbf{x}, \mathbf{r}) definiert in der Form

$$\mathbf{X} = \left\{ \mathbf{x}^t, \mathbf{r}^t \right\}_{t=1}^N , \quad (8.3)$$

wobei t der Index für verschiedene Beispiele der Menge ist. Er kann sowohl für eine Ordnung und damit eine zusammenhängende Abfolge der geordneten Paare stehen als auch für einen nicht ordnenden Parameter; hierbei ist jede Permutation der Menge gleichaussagekräftig. Abbildung 8.1 zeigt das Training der Modellparameter. Die funktionelle Relation beschreibt den Zusammenhang der geordneten Paare ohne eine

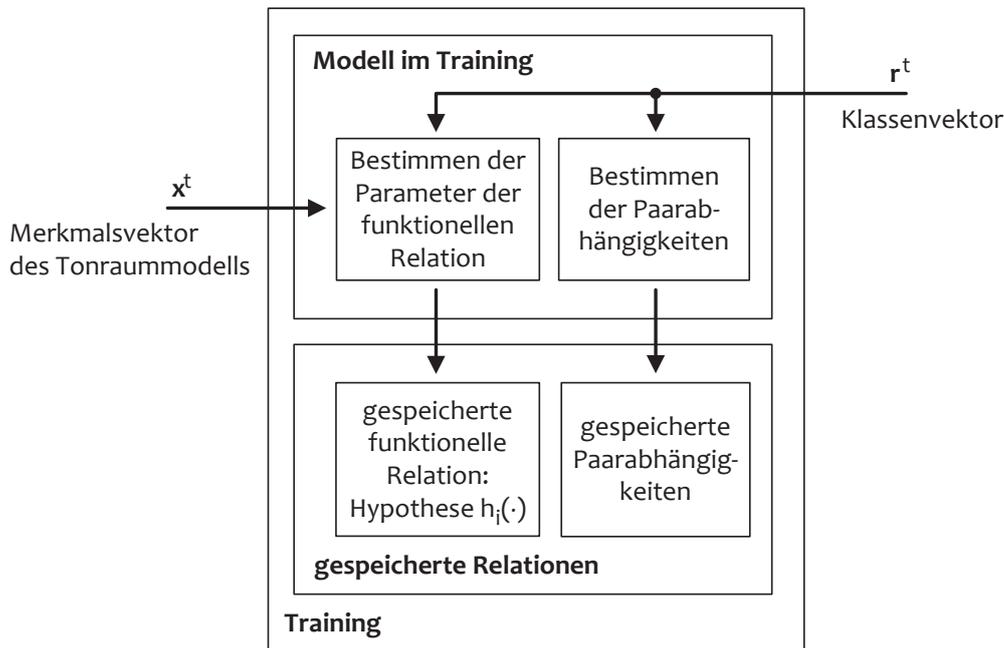


Abbildung 8.1 – Überwachtes Training mit geordneten Paaren

abhängige Reihenfolge mit K unterschiedlichen Hypothesen h_i , die durch Training die folgende Eigenschaft erlangen oder möglichst nahe an diese Eigenschaft herankommen sollen:¹

$$h_i(\mathbf{x}^t) = \begin{cases} 1 & \text{falls } \mathbf{x}^t \in C_i, \\ 0 & \text{falls } \mathbf{x}^t \in C_j, j \neq i. \end{cases} \quad (8.4)$$

Ein Modell² $g(\cdot)$ definiert die Hypothesenklasse, die alle Hypothesen $h_i(\cdot)$ zusammenfasst. Das Modell wird durch den Anwender festgelegt und bestimmt maßgeblich den Trainingserfolg. Es sollte in seiner Komplexität der Eingabedaten zugrundeliegenden Funktion entsprechen (vgl. [8], S.36). Wenn die Hypothese weniger komplex als die Funktion ist, so entsteht eine sogenannte *Unteranpassung*, im Gegenzug eine sogenannte *Überanpassung*. Bei der Modellauswahl müssen durch den Anwender Annahmen getroffen werden, die zu einer sogenannten *induktiven Verzerrung* führen können, da durch eine ungünstige

¹ Es steht nicht fest, dass nach dem Training die Eingabedaten den Klassen der Übungsmenge vom trainierten Modell vollständig richtig zugeordnet werden. Dies kann durch sogenannte *Anomalien* in den Daten hervorgerufen werden. Dazu zählen Ungenauigkeiten beim Aufzeichnen der Eingabedaten, Fehler bei den Klassenzuordnungen (engl.: *teacher noise*) oder ein unterangepasstes Modell. Vgl. [8], S.29.

² Als Modell können sowohl parametrische Modelle, wie z.B. eine Gaußverteilung oder *Support Vector Machines*, als auch nichtparametrische Modelle, wie z.B. *k Nearest Neighbours*, verwendet werden.

Modellwahl die Parameter der Hypothese nicht zur nachzubildenden Funktion passen.³ θ fasst alle Parameter der Hypothesen zusammen und instanziiert die Hypothesen des Modells $g(\cdot)$ in der Form

$$g(x|\theta) . \quad (8.5)$$

Um das Trainieren zu ermöglichen, ist eine *Verlustfunktion* $L(\cdot)$ zu definieren, welche die Differenz zwischen der Approximation, $g(x^t|\theta)$, und der gewünschten Ausgabe r^t berechnet. Der *Näherungsfehler* $E(\cdot)$ oder auch *Gesamtverlust* ist die Summe der Verluste der einzelnen Instanzen:

$$E(\theta|\mathbf{X}) = \sum_t L(r^t, g(x^t|\theta)) . \quad (8.6)$$

Beim Training selbst ist derjenige Wert θ^* zu finden, der den Näherungsfehler minimiert:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} E(\theta|\mathbf{X}) . \quad (8.7)$$

Manche Modelle, wie z.B. *Hidden Markov Models*, werten zusätzlich zu der funktionellen Relation auch die Abhängigkeiten der geordneten Paare aus, wenn solche vorhanden sind. Dies trifft in der Regel zu bei Zeitreihenanalysen, wie z.B. bei der Spracherkennung, der Wettervorhersage oder der Musikanalyse, sowie örtlich zusammenhängenden Folgen, z.B. der DNA-Analyse oder der Bildauswertung. Die Verwendung der Paarabhängigkeiten wird in Abschnitt 8.2.2.1 näher erläutert.

8.1.2 Klassifizierung mittels Bayesscher Entscheidungstheorie

Beim Testen (Klassifizieren) neuer Eingabedaten \mathbf{x} stellt sich die Frage, zu welcher Klasse C jeweils die einzelnen Eingabedaten gehören. Dazu muss die bedingte Wahrscheinlichkeit $P(C|\mathbf{x})$ berechnet werden. Dazu wird der Satz von Bayes ([8], S.46) zu Hilfe genommen:

$$P(C|\mathbf{x}) = \frac{P(C)p(\mathbf{x}|C)}{p(\mathbf{x})} . \quad (8.8)$$

$P(C=1)$ ist dabei die sogenannte *a-priori-Wahrscheinlichkeit* dafür, dass die Klasse C auftritt. Das ist das Wissen, welches durch das Lernen entstanden ist. Dieses Wissen wurde also vor dem Testen und damit *a-priori* erlangt. Es gilt weiterhin:

$$P(C = 0) + P(C = 1) = 1 . \quad (8.9)$$

³ Wenn keine Annahmen zu treffen wären, dann wüsste der Anwender alles über die Funktion und müsste folglich nicht trainieren, da die Lösung bekannt wäre.

$p(\mathbf{x}|C)$ ist die sogenannte *Klassen-Likelihood* und die bedingte Wahrscheinlichkeit dafür, dass \mathbf{x} bei gegebener Klasse C auftritt. $p(\mathbf{x})$ wird *Evidenz* genannt und drückt aus, mit welcher Wahrscheinlichkeit überhaupt die Beobachtung \mathbf{x} gemacht wird. $P(C|\mathbf{x})$ ist die sogenannte *a-posteriori-Wahrscheinlichkeit*. Damit kann Gleichung 8.8 wie folgt vereinfacht geschrieben werden:

$$\text{a-posteriori-Wahrscheinlichkeit} = \frac{\text{a-priori-Wahrscheinlichkeit} \times \text{Klassen-Likelihood}}{\text{Evidenz}} \quad (8.10)$$

Im allgemeinen Fall gibt es K sich gegenseitig ausschließende Klassen $C_i, i = 1, \dots, K$. Für die a-priori-Wahrscheinlichkeiten der Klassen muss Folgendes gelten:

$$P(C_i) \geq 0 \text{ und } \sum_{i=1}^K P(C_i) = 1 \quad (8.11)$$

Die a-posteriori-Wahrscheinlichkeit der Klasse C_i kann nun berechnet werden als:

$$P(C_i|\mathbf{x}) = \frac{P(C_i)p(\mathbf{x}|C_i)}{p(\mathbf{x})} = \frac{P(C_i)p(\mathbf{x}|C_i)}{\sum_{k=1}^K p(\mathbf{x}|C_k)P(C_k)} \quad (8.12)$$

Der *Bayessche Klassifikator* wählt die Klasse mit der höchsten a-posteriori-Wahrscheinlichkeit aus:

$$\text{wähle } C_i \text{ falls } P(C_i|\mathbf{x}) = \underset{n}{\operatorname{argmax}} P(C_n|\mathbf{x}) \quad (8.13)$$

Damit ist die Klassifizierung abgeschlossen.

Es gibt Anwendungsfälle, bei denen eine falsche Entscheidung zu treffen für unterschiedliche Klassen unterschiedlich schwer wiegt.⁴ Hierbei wird eine Verlustfunktion definiert, mit der ein zu erwartendes Risiko berechnet werden kann. Bei solchen Anwendungen wird dann das zu erwartende Risiko minimiert. Wenn jedoch der Verlust bei einer Fehlklassifizierung für alle Klassen gleich groß ist, dann ergibt sich als Klassifizierung automatisch die Ermittlung der wahrscheinlichsten Klasse mittels Bayesschem Klassifikator. In dieser Arbeit wird davon ausgegangen, dass alle Verluste gleichgewichtig sind.

8.2 Verwendete Maschinelle Lernverfahren

Dieser Abschnitt stellt zuerst die beiden Maschinenlernverfahren *k Nearest Neighbors* (KNN) und *Hidden Markov Models* (HMM) vor. Anschließend wird im dritten Abschnitt

⁴ Dies kann z.B. bei der Kreditvergabe eines Geldinstitutes der Fall sein. Hier geht es nicht nur darum, mit einem ausgezahlten Darlehen einen möglichst hohen Gewinn zu erzielen, sondern auch abzuwägen, ob der Kunde vertrauenswürdig ist und das Darlehen überhaupt zurückzahlen kann.

ein für die Akkorderkennung neues Verfahren eingeführt, das HMM mit einem KNN-Schätzer kombiniert. Die Leistungsfähigkeit des neuen Verfahrens wird in Kapitel 12 mit der Leistungsfähigkeit des Verfahrens HMM und des Verfahrens KNN verglichen.

8.2.1 KNN – k Nearest Neighbors

KNN gehört zu den instanzbasierten Verfahren und ist ein sogenannter *fauler Lerner* (engl.: lazy learner; vgl. [115], S.230). Im Gegensatz zu den parametrischen Verfahren, die während des Trainings „eifrig“ ihre Hypothesen anpassen, speichern einfache instanzbasierte Verfahren während des Trainings lediglich die Übungsmenge X ohne zu rechnen oder irgendetwas anzupassen. Damit haben instanzbasierte Verfahren erstens einen wesentlich höheren Speicherbedarf als parametrische Verfahren und verschieben zweitens den Rechenaufwand in den Klassifizierungsvorgang (vgl. [8], S.162). Abbildung 8.2 zeigt ein einfaches Beispiel, bei dem die Eingabedaten x zweidimensional sind und pro Klasse C_i sieben Instanzen gespeichert wurden. Insgesamt gibt es zwei unterschiedliche Klassen. Vor der Klassifizierung wird festgelegt, wieviele am nächsten liegende „Nachbarn“ (Trainingsdatenpunkte) k gesucht werden sollen, um die Klassenzugehörigkeit des zu klassifizierenden Elements zu ermitteln.

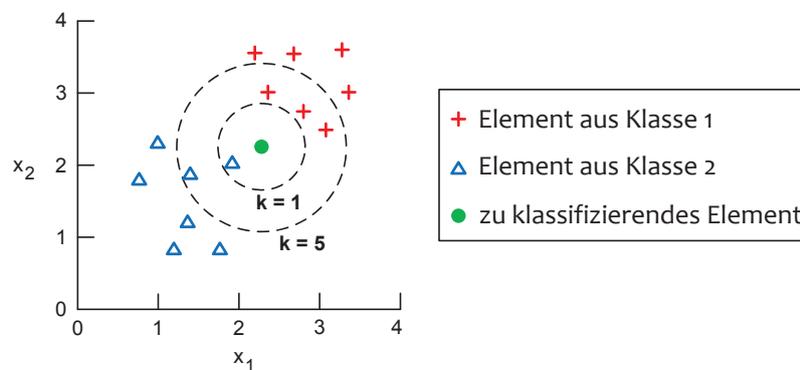


Abbildung 8.2 – Klassifizierung mit KNN

Um diejenigen Nachbarn zu finden, die am nächsten liegen, muss ein Abstandsmaß (Metrik) $d(\cdot)$ gewählt werden. Dies kann einem beliebigen metrischen Raum entstammen. Für das gezeigte Beispiel wird die euklidische Distanz verwendet. Es wird nun diejenige Klasse C_i für das zu klassifizierende Element gewählt, deren Anzahl an Elementen unter den k nächsten Nachbarn am größten ist. Dies wird hier *einfaches Auszählen* genannt. Bei Gleichstand wird zufällig ausgewählt. Oft wird k ungerade gewählt, um die Chance für einen Gleichstand zu verringern. Neben der einfachen Anzahl an Nachbarn pro Klasse wird häufig das inverse Abstandskadrat als Zählmaß gewählt, was in dieser Arbeit als

gewichtetes Auszählen bezeichnet wird, um näher liegende Nachbarn zu bevorzugen und weiter weg liegende Nachbarn zu „bestrafen“. Dies wird im folgenden Abschnitt näher erläutert.

8.2.1.1 Klassifizierung bei KNN

Die Klassifizierung findet mittels Bayes-Schätzer statt:

$$\hat{P}(C_i|\mathbf{x}) = \frac{\hat{p}(\mathbf{x}|C_i)\hat{P}(C_i)}{\hat{p}(\mathbf{x})} . \quad (8.14)$$

Dazu müssen die a-priori-Wahrscheinlichkeit, die Klassen-Likelihood und die Evidenz beschrieben werden. Dies wird mit Hilfe von Verteilungsdichten durchgeführt ([8], S.168).

Einfaches Auszählen

Zuerst wird der Schätzer für die einfache Anzahl an Nachbarn hergeleitet. Die Evidenz stellt sich hierbei wie folgt dar:

$$\hat{p}(\mathbf{x}) = \frac{k}{N \cdot V^k(\mathbf{x})} , \quad (8.15)$$

wobei N die Gesamtanzahl an gespeicherten Trainingspunkten aller Klassen ist und $V^k(\mathbf{x})$ das Volumen einer aufgespannten Hypersphäre, bei der \mathbf{x} den Mittelpunkt markiert und genau k Trainingspunkte in der Hypersphäre enthalten sind. Dies beschreibt die Verteilungsdichte der k nächsten Nachbarn in Bezug auf alle Trainingspunkte der Übungsmenge X in dem Volumen. Die a-priori-Wahrscheinlichkeit $\hat{P}(C_i)$ ist das Verhältnis aus Anzahl der Trainingspunkte pro Klasse N_i und Gesamtanzahl aller gespeicherten Trainingspunkte:

$$\hat{P}(C_i) = \frac{N_i}{N} . \quad (8.16)$$

Die Klassen-Likelihood ergibt sich aus der Verteilungsdichte der Anzahl der Punkte k_i der betrachteten Klasse C_i , die im Volumen $V^k(\mathbf{x})$ liegen, und der Gesamtanzahl an Trainingspunkten N_i der betrachteten Klasse:

$$\hat{p}(\mathbf{x}|C_i) = \frac{k_i}{N_i \cdot V^k(\mathbf{x})} . \quad (8.17)$$

Durch Einsetzen von Gleichung 8.15, Gleichung 8.16 und Gleichung 8.17 in Gleichung 8.14 kann der Schätzer für das einfache Auszählen hergeleitet werden (vgl. [8], S.170):

$$\hat{P}(C_i|\mathbf{x}) = \frac{\hat{p}(\mathbf{x}|C_i)\hat{P}(C_i)}{\hat{p}(\mathbf{x})} = \frac{\frac{k_i}{N_i \cdot V^k(\mathbf{x})} \cdot \frac{N_i}{N}}{\frac{k}{N \cdot V^k(\mathbf{x})}} = \frac{k_i}{k} . \quad (8.18)$$

Es wird diejenige Klasse C_i ausgewählt, welche die größte Anzahl an Trainingspunkten k_i bei gegebenem \mathbf{x} unter den k nächsten Nachbarn hat:

$$\text{wähle } C_i \text{ falls } \hat{P}(C_i|\mathbf{x}) = \operatorname{argmax}_n \frac{k_n}{k} . \quad (8.19)$$

Gewichtetes Auszählen

Bei Verwendung des inversen Abstandsquadrates

$$\frac{1}{d(\mathbf{x})^2} \quad (8.20)$$

als Zählmaß sind für die Dichteschätzungen die Anzahlen k und k_i durch die Summen der inversen Abstandsquadrate zu ersetzen. Damit wird k ersetzt wie folgt:

$$k \leftarrow \sum_{s=1}^k \frac{1}{d(\mathbf{x}, \mathbf{x}_{V^k}^s)^2} , \quad (8.21)$$

wobei $\mathbf{x}_{V^k}^s$ alle k nächsten Nachbarn im Volumen $V^k(\mathbf{x})$ sind. k_i wird auf die gleiche Weise ersetzt:

$$k_i \leftarrow \sum_{s=1}^{k_i} \frac{1}{d(\mathbf{x}, \mathbf{x}_{V_i^k}^s)^2} , \quad (8.22)$$

wobei hier $\mathbf{x}_{V_i^k}^s$ diejenigen k_i nächsten Nachbarn der Klasse C_i im Volumen $V^k(\mathbf{x})$ sind. Der Schätzer für das gewichtete Auszählen ergibt sich durch Einsetzen von Gleichung 8.21 und Gleichung 8.22 in Gleichung 8.18:

$$\hat{P}(C_i|\mathbf{x}) = \frac{\sum_{s=1}^{k_i} \frac{1}{d(\mathbf{x}, \mathbf{x}_{V_i^k}^s)^2}}{\sum_{s=1}^k \frac{1}{d(\mathbf{x}, \mathbf{x}_{V^k}^s)^2}} . \quad (8.23)$$

Es gilt:

$$\sum_{i=1}^K \hat{P}(C_i|\mathbf{x}) = \sum_{i=1}^K \frac{\sum_{s=1}^{k_i} \frac{1}{d(\mathbf{x}, \mathbf{x}_{V_i^k}^s)^2}}{\sum_{s=1}^k \frac{1}{d(\mathbf{x}, \mathbf{x}_{V^k}^s)^2}} = 1 . \quad (8.24)$$

Es wird diejenige Klasse C_i bei gegebenem \mathbf{x} ausgewählt, die Gleichung 8.23 maximiert:

$$\text{wähle } C_i \text{ falls } \hat{P}(C_i|\mathbf{x}) = \operatorname{argmax}_n \frac{\sum_{s=1}^{k_n} \frac{1}{d(\mathbf{x}, \mathbf{x}_{V_n^k}^s)^2}}{\sum_{s=1}^k \frac{1}{d(\mathbf{x}, \mathbf{x}_{V^k}^s)^2}} . \quad (8.25)$$

8.2.2 HMM – Hidden Markov Models

HMM verwenden sowohl die funktionelle Relation als auch die Paarabhängigkeiten (siehe Abbildung 8.1). Im Folgenden werden zuerst die Eigenschaften von Markov-Ketten erklärt. Anschließend werden HMM in ihren Parametern erläutert und die verschiedenen Anwendungsfälle aufgezeigt.

8.2.2.1 Markov-Ketten

Eine Markov-Kette⁵ ist ein Abfolge von definierten Zuständen und den damit verbundenen Zustandsänderungen, die mit Übergangswahrscheinlichkeiten beschreibbar sind. Eine Markov-Kette bildet in den vorgegebenen Grenzen ein Modell, das in der Lage ist vorherzusagen, welcher Zustand auf einen anderen mit welcher Wahrscheinlichkeit folgt. Als Beispiel (vgl. [129], S.321 ff.) wird eine Zustandsmenge definiert $\{\text{sonnig}, \text{bewölkt}, \text{regnerisch}\}$, die geeignet ist, das Wetter zu beschreiben. Diese Menge besteht aus $N = 3$ Zuständen.

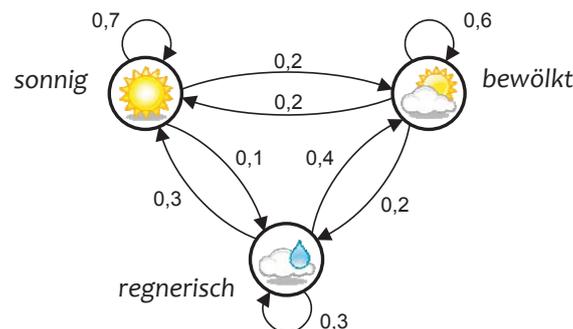


Abbildung 8.3 – Beispiel für eine Markov-Kette

Es ist ein Zeitintervall festzulegen, nach dem immer periodisch der Zustand des betrachteten Systems nachgefragt bzw. später in der Analyse geschätzt wird. Das Zeitintervall wird hier im Beispiel auf genau einen Tag festgelegt, d.h. es wird jeden Tag einmal überprüft, welchen Zustand das Wetter hat. Abbildung 8.3 zeigt das Beispiel mit den drei Zuständen *sonnig*, *bewölkt* und *regnerisch*. Die Übergangswahrscheinlichkeiten zwischen den Zuständen sind hier bereits vorgegeben. Die gezeigte Markov-Kette drückt aus, dass z.B. auf einen Sonnentag mit 70 % Wahrscheinlichkeit wieder ein Sonnentag folgt⁶, und dass auf einen regnerischen Tag mit 40 % Wahrscheinlichkeit ein bewölkter Tag folgt.

⁵ Der Begriff geht auf den russischen Mathematiker Andrei Andrejewitsch Markow (1856-1922) zurück.

⁶ Dies ist am Pfeil, der vom Zustand *sonnig* auf wieder *sonnig* verweist, und an der Zahl 0,7 zu erkennen.

Die Übergangswahrscheinlichkeiten können durch Beobachtung des Wetters bestimmt werden. Dies ist mit dem Training der Markov-Kette gleichzusetzen. Nach diesem einführnden Beispiel folgt nun eine allgemeine Beschreibung von Markov-Ketten.

Bei einer Markov-Kette werden zu den Zeitpunkten $t = 1, 2, \dots, T$ die Zustände q_t ermittelt. Die Wahrscheinlichkeit, dass ein Zustand q_t genau zum Zeitpunkt t auftritt, hängt von allen Vorgängerzuständen ab. Zur Vereinfachung wird eine sogenannte *Markov-Kette erster Ordnung* definiert (vgl. [129], S.323). Ein Zustand q_t hängt dadurch nur vom direkten Vorgänger ab:

$$P(q_t = j | q_{t-1} = i, q_{t-2} = h, \dots) = P(q_t = j | q_{t-1} = i). \quad (8.26)$$

Des Weiteren wird die Markov-Kette als zeitinvariant definiert, d.h. auf einen Zustand $q_t = i$ folgt der Zustand $q_{t+1} = j$ immer mit der gleichen Wahrscheinlichkeit, unabhängig vom Auftrittszeitpunkt des Zustandes $q_t = i$. Dies führt zu den Übergangswahrscheinlichkeiten a_{ij} :

$$a_{ij} = P(q_t = j | q_{t-1} = i) \quad \text{mit} \quad 1 \leq i, j \leq N. \quad (8.27)$$

mit den folgenden Eigenschaften

$$a_{ij} \geq 0 \quad \forall j, i \text{ und} \quad (8.28)$$

$$\sum_{j=1}^N a_{ij} = 1 \quad \forall i. \quad (8.29)$$

Damit ist eine Markov-Kette erster Ordnung vollständig beschrieben.

Die Übergangswahrscheinlichkeiten a_{ij} lassen sich in der Zustandsübergangsmatrix A kompakt zusammenfassen. Für das zuvor genannte Beispiel mit den drei unterschiedlichen Wetterzuständen ergibt sich:

$$A = \{a_{ij}\} = \begin{bmatrix} 0,7 & 0,2 & 0,1 \\ 0,2 & 0,6 & 0,2 \\ 0,3 & 0,4 & 0,3 \end{bmatrix} \quad (8.30)$$

8.2.2.2 Modellparameter von HMM

Die Grundlage für HMM bildet die zuvor beschriebene Markov-Kette mit ihren Zuständen. In dieser Arbeit werden Markov-Ketten, die in der Zeit existieren, betrachtet, und die Paarabhängigkeiten werden hier *zeitliche Relation* genannt. Die Zustände, die später zu bestimmenden Akkorde, sind bei HMM nicht direkt sichtbar (engl.: *hidden*), sondern

nur über einen stochastischen Prozess schätzbar. Der stochastische Prozess erzeugt in Abhängigkeit der tatsächlichen Zustände eine Beobachtungssequenz (Beobachtungsmatrix)

$$\mathbf{O} = \{\mathbf{o}_1, \mathbf{o}_1, \dots, \mathbf{o}_T\}, \quad (8.31)$$

wobei die Beobachtungsvektoren \mathbf{o}_t mit $1 \leq t \leq T$ in dieser Arbeit den Merkmalsvektoren des ausgewählten Tonraummodells entsprechen. Während der Testphase (Klassifizierung) wird nicht jeder Merkmalsvektor sofort einer Klasse zugeordnet, sondern mit jedem Merkmalsvektor wird zuerst ermittelt, wie wahrscheinlich zum betrachteten Zeitpunkt alle einzelnen Zustände sind. Erst danach findet die Zuordnung zu den Klassen statt, indem mit der zeitlichen Relation die wahrscheinlichste Zustandsfolge ermittelt wird. Dies erfolgt mit Hilfe des sogenannten Viterbi-Algorithmus.⁷

Ein Hidden Markov Model wird durch folgende Parameter vollständig beschrieben:

1. Anzahl N der Zustände (Klassen)
2. Anzahl M der Dimensionen der Beobachtungsvektoren (Merkmalsvektoren) mit dazugehörigem Vektor:

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_M \end{bmatrix} \quad (8.32)$$

3. Zustandsübergangsmatrix (zeitliche Relation) $A = \{a_{ij}\}$ mit

$$a_{ij} = P(q_{t+1} = j | q_t = i) \quad 1 \leq i, j \leq N \quad (8.33)$$

4. Wahrscheinlichkeitsschätzer (funktionelle Relation) $\mathbf{b} = \{b_j\}$ $1 \leq j \leq N$ mit

$$b_j = P(\mathbf{o}_t = \mathbf{m} | q_t = j) \quad (8.34)$$

5. Anfangsverteilung π mit

$$\pi_i = P(q_1 = i) \quad 1 \leq i \leq N \quad (8.35)$$

Die Anfangsverteilung π gibt an, wie wahrscheinlich jeder Zustand zu Beginn einer Beobachtungssequenz ist. Diese verwendet der Viterbi-Algorithmus als Initialisierung,

⁷ Vgl. Abschnitt 8.2.2.4.

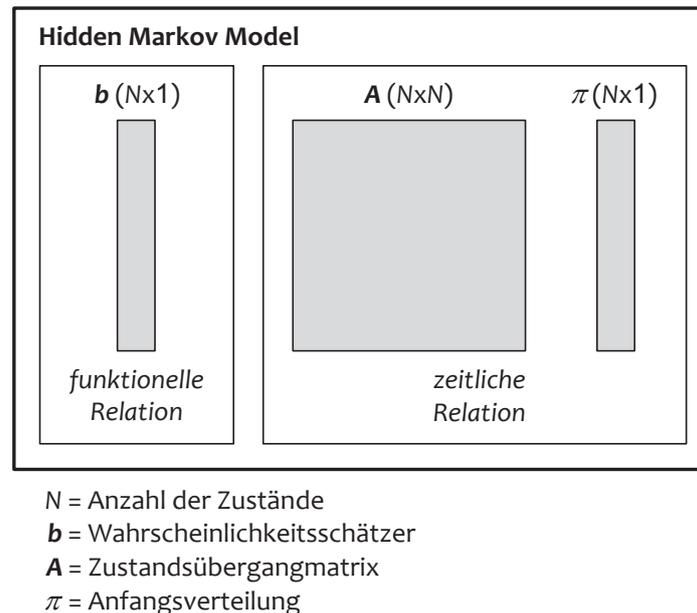


Abbildung 8.4 – HMM Modellparameter

um die Klassifizierung vornehmen zu können. Abbildung 8.4 fasst die Parameter eines HMM zusammen. Es wird folgende kompakte Schreibweise für ein HMM definiert:

$$\lambda = \{A, b, \pi\} \quad (8.36)$$

Die Parameter N und M sind im Modell automatisch mit eingeschlossen.

8.2.2.3 HMM-Anwendungsfälle

Es gibt für HMM drei typische Anwendungsfälle (siehe Abbildung 8.5). Bei allen wird davon ausgegangen, dass ein bereits trainiertes HMM λ vorliegt. Zudem ist die Beobachtungsmatrix O gegeben.

Fall 1 – Evaluierung

Bei der Evaluierung existieren mehrere HMM, wie z.B. bei der Spracherkennung. Es ist $P(O|\lambda)$ für alle Modelle zu berechnen. Dies erfolgt mit der sogenannten *Forward*-Prozedur. Alternativ gibt es hierfür auch die sogenannte *Backward*-Prozedur. Eine genaue Beschreibung ist in [129] zu finden.

Fall 2 – Dekodierung

Der Anwendungsfall Dekodierung (Analysephase) ist für diese Arbeit interessant und wird in Abschnitt 8.2.2.4 näher erläutert. Es ist für ein trainiertes HMM und eine gegebene Beobachtungsmatrix die wahrscheinlichste Zustandsfolge zu finden.

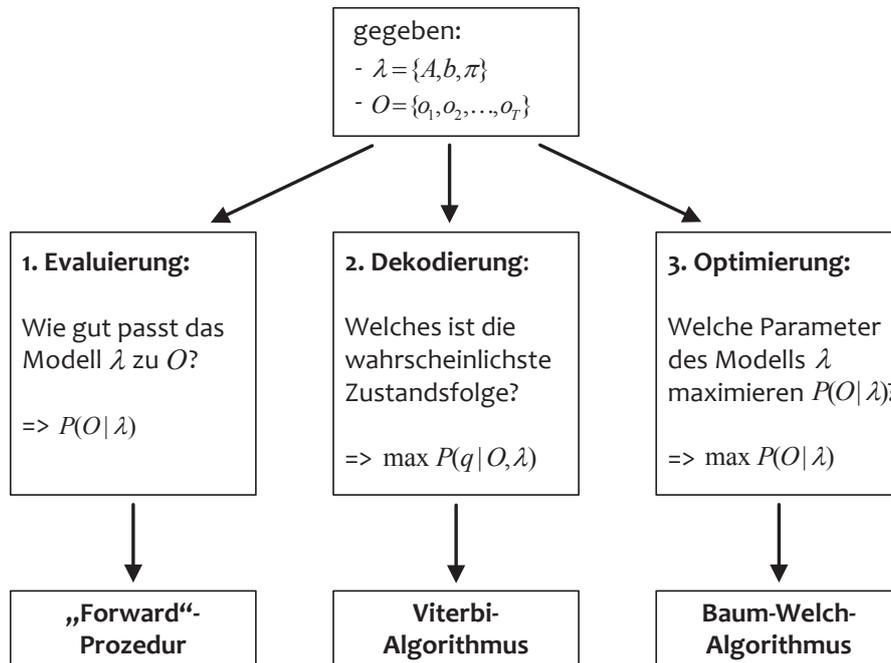


Abbildung 8.5 – HMM – typische Anwendungsfälle und Berechnungsmethoden

Fall 3 – Optimierung

Bei der Optimierung werden die Parameter A , b und π eines bereits (z.B. durch Expertenwissen) „vor“-trainierten HMM angepasst. Dafür wird der sogenannte Baum-Welch-Algorithmus verwendet. Die Parameteranpassung ist dann notwendig, wenn nur eine Übungsmenge X mit Eingabedaten vorliegt, jedoch keine geordneten Paare. Ausführliche Beschreibungen dazu sind in [129] zu finden.

8.2.2.4 Dekodierung – Klassifizierung bei HMM

Die Dekodierung wird in zwei Schritten durchgeführt. Im ersten Schritt wird aus der Beobachtungsmatrix O die Zustandswahrscheinlichkeitsmatrix B mit Hilfe von b berechnet, indem die Beobachtungsvektoren o_t wie in Gleichung 8.34 beschrieben auf die Matrix B abgebildet werden. Die Elemente (funktionelle Relationen zu den Klassen) des Wahrscheinlichkeitsschätzers b sind in der Regel einzeln trainierte Gaußverteilungen oder einzeln trainierte Gaußmischverteilungen (vgl. [8], S.91). B ist eine Matrix, die zu jedem Zeitpunkt t angibt, wie wahrscheinlich welcher Zustand ist. Im zweiten Schritt ermittelt der Viterbi-Algorithmus den wahrscheinlichsten Pfad durch die Matrix B . Dieser Vorgang ist die Klassifizierung; die Zustände entlang des Pfades sind die ermittelten Klassen.

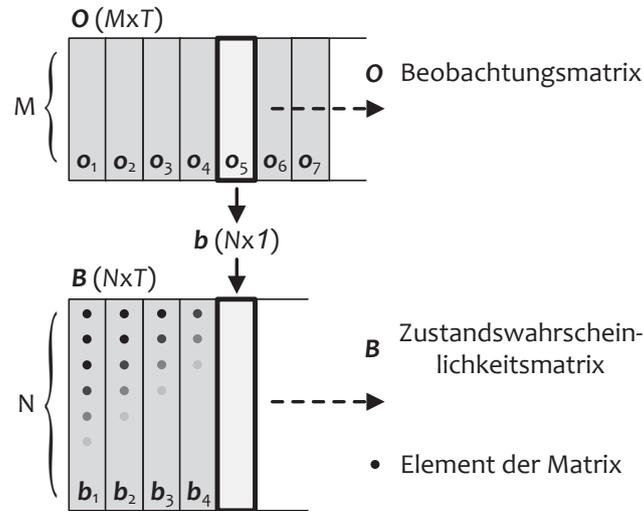


Abbildung 8.6 – Berechnen der Zustandswahrscheinlichkeitsmatrix B

Der **Viterbi-Algorithmus** berechnet zuerst jeweils den wahrscheinlichsten Vorgängerzustand zu allen Zuständen in der Matrix B :

$$\arg \max_{1 \leq i \leq N} \{P(q_{t-1} = i | q_t = j) \cdot a_{ij}\} \quad \forall j \text{ mit } 1 \leq j \leq N \quad (8.37)$$

Die Wahrscheinlichkeit, dass ein bestimmter Zustand auf einen anderen folgt, ist das Produkt der beiden Zustandswahrscheinlichkeiten $P(q_t = j)$ und $P(q_{t-1} = i)$ multipliziert mit der Übergangswahrscheinlichkeit a_{ij} zwischen den beiden Zuständen.

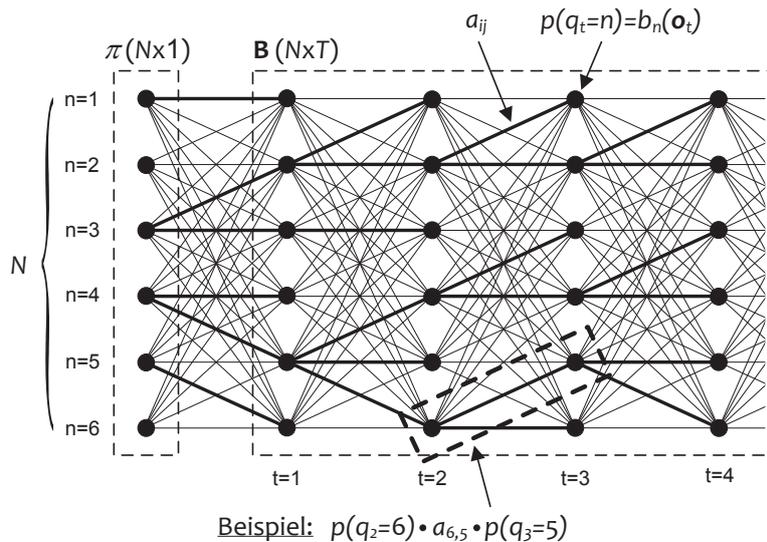


Abbildung 8.7 – Trellis-Diagramm bei der Viterbi-Dekodierung

Das Ergebnis der „bester Vorgängerzustand“-Berechnung ist im Beispiel in Abbildung 8.7 in der sogenannten *Trellis*- oder auch *Lattice*-Struktur zu sehen. Die hervorgehobenen schwarzen Kanten stellen jeweils die wahrscheinlichsten Übergänge mit den besten Vorgängerzuständen dar. Die Anfangsverteilung π ist der Vorgänger zum ersten Vektor in Matrix B und wird damit zur Initialisierung des Viterbi-Algorithmus verwendet.

Anschließend erfolgt die Klassifizierung mit dem sogenannten *Backtracking*. Es wird von hinten beginnend entlang der hervorgehobenen schwarzen Kanten der wahrscheinlichste Pfad durch das Trellis gesucht. Dazu sind entlang eines Pfades alle Wahrscheinlichkeiten zu multiplizieren (Zustandswahrscheinlichkeiten und Übergangswahrscheinlichkeiten). Es gibt maximal N Pfade. In der Regel reduziert sich die Anzahl der Pfade beim Weg durch das Trellis relativ schnell, da sich die Pfade an einigen Punkten treffen. Jedes Mal, wenn sich zwei Pfade treffen, scheidet einer der beiden Pfade aus, und zwar derjenige, der die kleinere akkumulierte Wahrscheinlichkeit hat. Der letzte übriggebliebene Pfad ist das Ergebnis der Klassifizierung. Abbildung 8.8 zeigt zusammenfassend den Vorgang der Dekodierung bei HMM als Blockschaltbild.

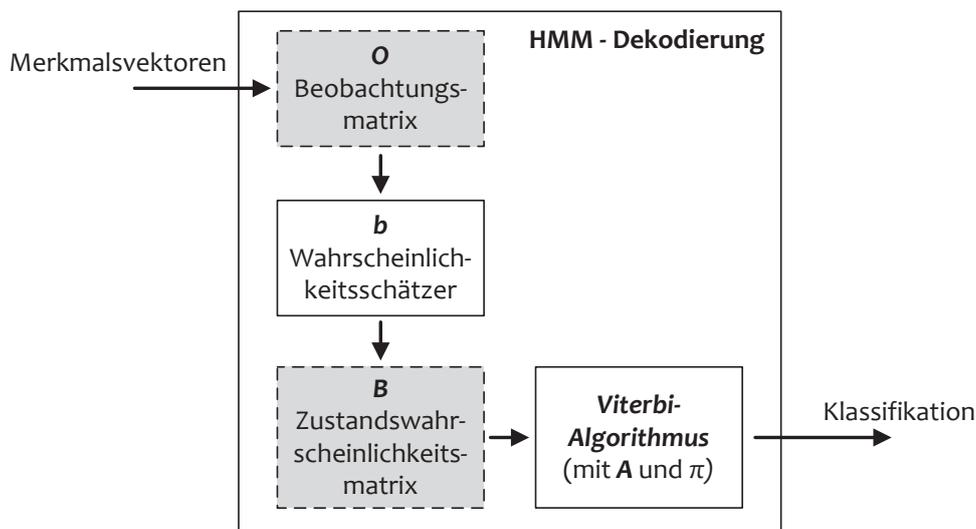


Abbildung 8.8 – HMM - Dekodierung

8.2.3 HMM mit KNN-Schätzer

Als Wahrscheinlichkeitsschätzer b eignet sich neben Gaußverteilungen auch der KNN-Schätzer aus Abschnitt 8.2.1. Lefèvre, Montacié und Caraty haben 1997 die kombinierte Verwendung von HMM und KNN für die Spracherkennung vorgeschlagen (vgl. [96], [97], [98]) und deren Leistungsfähigkeit aufgezeigt. In der Musikanalyse wurde dieses

Verfahren bisher weder verwendet noch evaluiert und stellt damit ein für das MIR neues Verfahren dar. Da das *gewichtete Auszählen* sehr leistungsfähig ist, kommt es hierbei zum Einsatz.

8.2.3.1 Klassifizierung mit KNN-Schätzer

Der KNN-Schätzer aus Gleichung 8.23 wird als Wahrscheinlichkeitsschätzer \mathbf{b} aus Gleichung 8.34 verwendet in der Form

$$b_i = \hat{P}(C_i|\mathbf{o}) = \frac{\sum_{s=1}^{k_i} \frac{1}{d(\mathbf{o}, \mathbf{o}_{v_i^s}^s)^2}}{\sum_{s=1}^k \frac{1}{d(\mathbf{o}, \mathbf{o}_{v^s}^s)^2}} \quad (8.38)$$

In Kapitel 12 wird die Leistungsfähigkeit des Maschinenlernverfahrens HMM (mit Gaußverteilung als Wahrscheinlichkeitsschätzer) mit dem hier für das MIR neuen Maschinenlernverfahren verglichen. Das neue Verfahren ist robuster in der Erkennung und erzielt wesentlich bessere Ergebnisse.

Kapitel 9

Stand der Technik Musikanalyse

Das folgende Kapitel beschreibt den Stand der Technik in den Bereichen der Tonarterkennung und Akkorderkennung.

Die *Tonart* umfasst das sogenannte tonale Zentrum und den grundlegenden harmonischen Charakter eines Musikstücks oder einer Musikpassage. Ein Musikstück kann aus mehreren Passagen unterschiedlicher Tonart bestehen. Ein gleitender Übergang zwischen zwei Passagen unterschiedlicher Tonart wird mit dem Begriff „Modulation“ bezeichnet. Die *Akkorde* beschreiben in ihrer zeitlichen Abfolge die harmonische Entwicklung eines Musikstücks, wobei sich der Stand der Technik auf die Analyse von reinen Dur- und Moll-Dreiklängen beschränkt. Dazu werden die übrigen Akkorde – wie z.B. der Septakkord oder der Sixte ajoutée – auf den jeweiligen Dur- oder Moll-Dreiklang abgebildet.

Es folgt zuerst der Überblick für die Tonarterkennung. Danach wird der Stand der Technik für die Akkorderkennung dargestellt.

9.1 Tonarterkennung

Dieser Abschnitt stellt den Stand der Technik zur Tonarterkennung dar. Die Tonart gibt den grundlegenden Charakter eines Musikstücks oder einer Musikpassage vor. Die beiden am häufigsten verwendeten Tonarten¹ sind die Kirchentonarten *Ionisch* und *Aeolisch*², die allgemein mit Dur und Moll bezeichnet werden. Die Bezeichnungen Dur und Moll sind allerdings nicht eindeutig, da es drei verschiedene Dur-Kirchentonarten (Ionisch, Lydisch und Mixolydisch) und drei verschiedene Moll-Kirchentonarten (Dorisch, Phrygisch und Aeolisch) gibt. Die aktuellen Forschungen und Veröffentlichungen zum Stand

¹ Dies trifft für Stücke aller Musikepochen zu, auch für die neuzeitliche Pop-Musik.

² Vgl. Abschnitt 2.6.4.

der Technik zeigen, dass diese sechs Tonarten auf Dur und Moll reduziert werden. Das liegt an den vorgestellten Ansätzen zum Stand der Technik, die für eine Tonarterkennung mit allen Kirchentonarten nur bedingt geeignet sind. Hierbei gehen jedoch wichtige Informationen über das Musikstück verloren.³

Die meisten Arbeiten verwenden sogenannte „Tonartprofile“. Diese stellen einen definierten Chromavektor dar, der unterschiedliche Amplituden in den Tonigkeiten enthält. Das nächste Unterkapitel stellt die wichtigsten Tonartprofile vor. Danach werden zunächst die Algorithmen behandelt, die mit MIDI-Daten (symbolisches Audio) arbeiten, und anschließend die Algorithmen, die mit PCM-Daten (reales Audio) arbeiten.

9.1.1 Tonartprofile

Ein Tonartprofil ist ein Chromavektormuster, das während der Analysephase mit zu analysierenden Merkmalsvektoren (notwendigerweise ebenfalls Chromavektoren) verglichen wird. Der Vergleich kann auf mehrere Arten durchgeführt werden, wobei im Prinzip immer ein Abstandsmaß verwendet wird.⁴ Die Tonartprofile existieren für jede Tonigkeit kombiniert mit Dur oder Moll. Eine Ausnahme bildet das diatonische Tonartprofil, das in Abbildung 9.1 zu sehen ist. Hierbei erhalten alle Tonigkeiten des Tonsets den Wert 1 und die restlichen tonsetfremden Tonigkeiten den Wert 0.

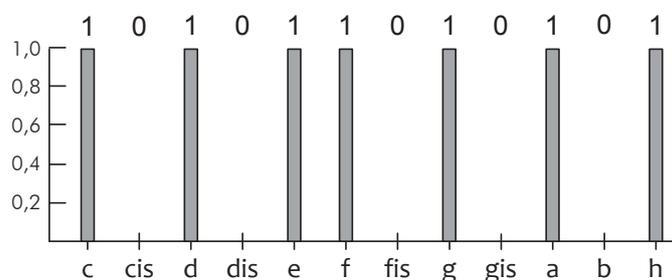


Abbildung 9.1 – Diatonisches Tonartprofil C-Dur

Dieses Profil ist damit nur geeignet, um ein bestimmtes Tonset von den anderen Tonsets zu unterscheiden. Es differenziert jedoch nicht zwischen Dur und Moll. Die Tonartprofile für die restlichen elf Tonsets werden durch einfache Rotation der Werte des Chromavektormusters generiert. Wenn z.B. alle Werte um eine Tonigkeit nach oben verschoben werden (der Wert von *h* wird dabei nach *c* abgebildet), dann wird aus dem Tonset von C-Dur das Tonset von Cis-Dur erzeugt. Das diatonische Tonartprofil für C-Dur ist wie

³ Vgl. Kapitel 11.

⁴ z.B. Korrelation, euklidische Distanz.

folgt definiert, wobei die erste Tonigkeit c ist:

$$P_D = [1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 1]$$

Krumhansl und Kessler (KK) haben im Jahr 1982 Experimente durchgeführt, um herauszufinden, wie gut die unterschiedlichen Tonigkeiten zu einer bestimmten Tonart passen (vgl. [82]). Dazu haben sie während der Untersuchungen zuerst mit Hilfe einer Kadenz⁵ einen musikalischen Kontext hergestellt. Nach einer kurzen Pause wurde jeweils eine der Tonigkeiten gespielt, und die Probanden sollten auf einer Skala angeben, wie gut diese Tonigkeit zu der davor gespielten Kadenz passt. Die Ergebnisse dieser Experimente sind in Abbildung 9.2 zu sehen, wobei nicht die absoluten Werte sondern die Relationen untereinander entscheidend sind.

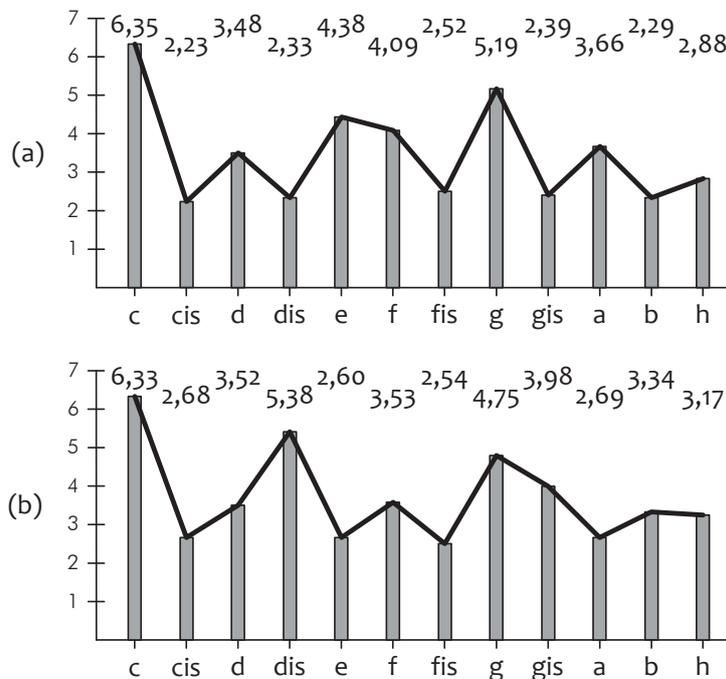


Abbildung 9.2 – Krumhansl/Kessler-Profile für (a) C-Dur und (b) c-Moll

Wie erwartet wird der Grundton einer Tonart dieser selbst am besten zugeordnet. Dies ist an den Werten 6,35 und 6,33 für Dur und Moll zu erkennen. Bei der Dur-Tonleiter folgen anschließend die Quinte und dann die Terz mit 5,19 und 4,38. Bei der Molltonleiter ist die Terz mit 5,38 nach dem Grundton die passendste Tonigkeit. Die tonartfremden Töne werden von den Probanden als am wenigsten dazugehörig empfunden. Für die unterschiedlichen Gewichtungen der einzelnen Tonigkeiten werden mehrere mögliche

⁵ Die Kadenz endet mit der Tonika.

Gründe genannt. Einer ist, dass die Obertöne eines natürlichen Tons nach der Oktave zuerst die Quinte und anschließend die Terz enthalten. Dies passt beim Dur-Tonartprofil. Eine andere Begründung ist, dass die drei Hauptfunktionen (Tonika, Subdominate und Dominante) zusammen eine Tonart festlegen und das Auftreten der einzelnen Tonigkeiten beim Spielen einer Kadenz unterschiedlich häufig ist. Die Kadenz in C-Dur enthält am häufigsten die Tonigkeiten *c* und *g*. Das Tonartprofil kann daher auch als Häufigkeitsprofil der einzelnen Tonigkeiten in der entsprechenden Tonart interpretiert werden. Es gibt einige Ansätze⁶, die aus den Trainingsstücken neue Tonartprofile erzeugen. Die Werte der Krumhansl/Kessler-Tonartprofile P_{KK_Dur} (Dur) und P_{KK_Moll} (Moll) sind in Tabelle 9.1 mit den anderen Profilen aufgelistet.

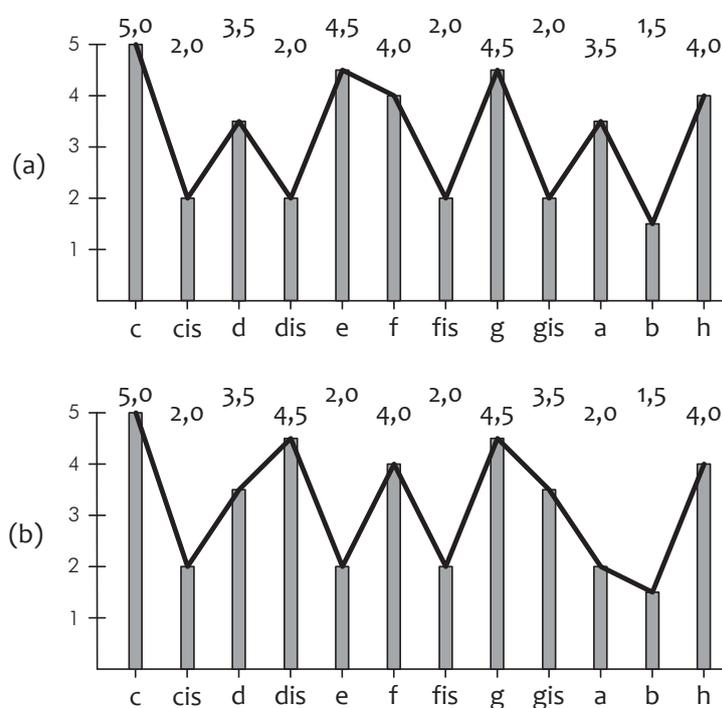


Abbildung 9.3 – Temperley-Profile für (a) C-Dur und (b) c-Moll

Im Gegensatz zu KK untersucht David Temperley (TY) Musik von der strukturellen und nicht von der wahrnehmungspsychologischen Seite (vgl. [158]). Er schlägt eigene Tonartprofile vor, welche die KK-Profile zur Grundlage haben und in Abbildung 9.3 zu sehen sind. Ein Nachteil der KK-Profile ist, dass die Moll-Tonart bevorzugt wird, da hierbei das Gesamtgewicht des Profils (Summe der zwölf Einzelgewichte) größer ist als beim Dur-Tonartprofil. Dadurch wird ein Musikstück wahrscheinlicher einer Moll-Tonart

⁶ Vgl. Abschnitt 9.1.3.

als einer Dur-Tonart zugeordnet. Es gibt aber bei der Analyse eines Musikstückes keinen Grund, dass eine der beiden Tongeschlechter (hier Moll) bevorzugt wird. Temperley wählt einfache Verhältnisse zwischen den verschiedenen Tonigkeiten, und die Summe der Gewichte ist in beiden Tonartprofilen gleich groß. Der Grundton des Tonartprofils wird nicht ganz so stark bewertet wie in den KK-Profilen. Des Weiteren werden bei beiden TY-Profilen die Terz und Quinte gleichwertig behandelt (4,5) und der Leitton (Halbton unter dem Grundton) erhält im Vergleich zu den KK-Profilen ein wesentlich höheres Gewicht (4,0). Die starke Gewichtung des Leittons bei Moll zeigt, dass für Temperley das harmonische Moll wichtig ist, da hier im Gegensatz zum natürlichen Moll der Leitton vorhanden ist.⁷ Das Dur-Tonartprofil P_{TY_Dur} von Temperley und das Moll-Tonartprofil P_{TY_Moll} werden zusammen mit den anderen Profilen in Tabelle 9.1 dargestellt.

	P_D	P_{KK_Dur}	P_{KK_Moll}	P_{TY_Dur}	P_{TY_Moll}
c	1	6,35	6,33	5,0	5,0
cis	0	2,23	2,68	2,0	2,0
d	1	3,48	3,52	3,5	3,5
dis	0	2,33	5,38	2,0	4,5
e	1	4,38	2,60	4,5	2,0
f	1	4,09	3,53	4,0	4,0
fis	0	2,52	2,54	2,0	2,0
g	1	5,19	4,75	4,5	4,5
gis	0	2,39	3,98	2,0	3,5
a	1	3,66	2,69	3,5	2,0
b	0	2,29	3,34	1,5	1,5
h	1	2,88	3,17	4,0	4,0

Tabelle 9.1 – Übersicht Tonartprofile für die Tonigkeit *c*

Es folgt der Stand der Technik zur Tonarterkennung mit symbolischem Audio und anschließend mit realem Audio.

⁷ Beim natürlichen Moll ist der Ton unter dem Grundton einen Ganzton entfernt und nicht wie der Leitton einen Halbtonschritt.

9.1.2 Tonarterkennung mit symbolischem Audio

Bei der Tonarterkennung mit symbolischem Audio wird nicht das Musiksignal direkt untersucht, sondern es wird versucht, aus dem Notenbild, das in Form von MIDI-Daten vorliegt, die Tonart des Musikstücks oder der Musikpassage zu bestimmen. Abbildung 9.4 zeigt den Ablauf der Analyse als Blockschaltbild.

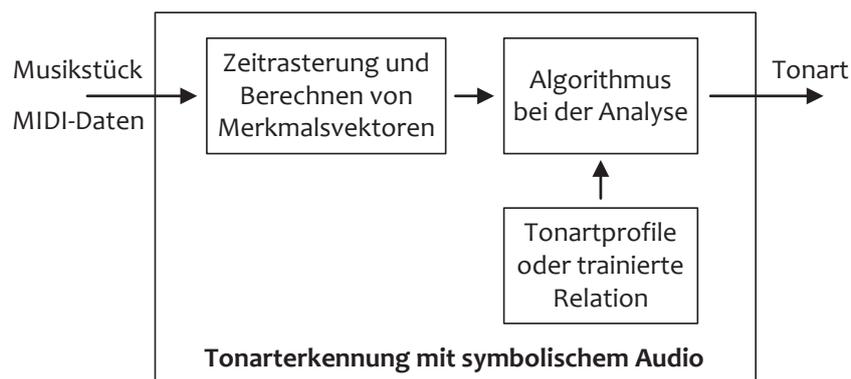


Abbildung 9.4 – Tonarterkennung mit symbolischem Audio

Madsen und Widmer verfolgen in ihrer Veröffentlichung „Key-Finding with Interval Profiles“ [104] die Idee, dass aus den aufeinanderfolgenden Intervallen einer Melodie auf die Tonart geschlossen werden kann. Dazu erstellen sie für jede Tonart ein Intervall-Profil (12 x 12 Matrix pro Tonart), das angibt, wie oft welches konsekutive Intervall auftritt. Diese Intervall-Profile werden mit MIDI-Melodien erlernt. Die anschließende Tonarterkennung beruht auf einfacher Korrelation der analysierten Intervalle mit den trainierten Intervall-Profilen. Der vorgestellte Algorithmus wird mit KK- und TY-Profilen verglichen und erzielt bei der Tonarterkennung von finnischen Folksongs die besten Ergebnisse, die im Bereich 60 - 80 % liegen.⁸ Bei der Analyse des *cantus firmus*⁹ von Bachfugen schneidet der dargestellte Algorithmus mit ca. 70 % schlechter ab als die KK- und TY-Profile. Die Evaluierung des Algorithmus zeigt zwei Schwachpunkte: Der Vergleich mit den KK- und TY-Profilen scheint nicht gerechtfertigt, da diese Profile für Musikstücke mit Akkordstrukturen erstellt worden sind. Hier werden jedoch Einzeltonfolgen analysiert. Außerdem sind finnische Folksongs eine ganz spezielle Musikrichtung, so dass die Ergebnisse daher nicht verallgemeinert werden können.

⁸ Die unterschiedlichen Ergebnisse wurden mit verschiedenen Vorverarbeitungen der Merkmalsvektoren erzielt.

⁹ Lat.: *cantus firmus* – die starke Stimme. In der Musik ist damit die melodieführende Stimme gemeint.

Noland und Sandler beschreiben in der Publikation „Key Estimation Using a Hidden Markov Model“ [139] einen HMM basierten Algorithmus, der pro Tonart ein HMM verwendet und in diesem die funktionalen Akkordfolgen speichert. Für das Training der HMM wird der EM-Algorithmus¹⁰ (Baum-Welch-Algorithmus) verwendet. Dazu wird eine Initialisierung der Matrix A und des Wahrscheinlichkeitsschätzers b benötigt. Hierfür werden die „correlations between key profiles“ und „harmonic hierarchy ratings for major, minor and diminished chords“ [82] von Carol Krumhansl verwendet. Die Anfangswahrscheinlichkeiten π werden alle zu $1/24$ gesetzt¹¹, um keine Tonart zu bevorzugen. Anschließend werden die drei Parameter der HMM in mehreren Durchgängen unterschiedlich kombiniert trainiert, um herauszufinden, für welchen Parameter das Training gut geeignet ist und für welchen es sich – auf das Analyseergebnis bezogen – negativ auswirkt. Es zeigt sich, dass das Training der Anfangswahrscheinlichkeiten keinen Einfluss auf das Ergebnis hat. Das Training des Wahrscheinlichkeitsschätzers b führt zu einer starken Verschlechterung der Ergebnisse, wobei jedoch das Training für die Übergangswahrscheinlichkeiten, Matrix A , eine Verbesserung der Ergebnisse erzielt. Für das Training und die Tests werden 110 Beatles-Stücke¹² verwendet. Das beste Ergebnis liegt bei 91 % richtig erkannter Tonarten. Es findet kein Vergleich zu anderen Verfahren statt. Die Evaluierung betrifft nur die Parametervariation im Training der HMM. Der vorgestellte Ansatz ist vielversprechend, da er die musiktheoretischen Funktionalbeziehungen in den Tonarten berücksichtigt. Dies wird aber nur ansatzweise bewusst im System realisiert, da die Akkordfolgen und nicht die Funktionen direkt trainiert werden. Für eine vollständige Funktionsanalyse müssten die Akkorde als Funktionen analysiert und vor allem die Stellvertreterklänge mit berücksichtigt werden, was hier nicht der Fall ist.

Elaine Chew hat ein eigenes Tonraummodell, das „Spiral Array“¹³, entwickelt, das geeignet sein soll, tonale Zentren zu modellieren und abzubilden. Das Modell wurde von ihr in mehreren Veröffentlichungen vorgestellt. In „The Spiral Array: An Algorithm for Determining Key Boundaries“ [22] wird das Tonraummodell verwendet, um Tonartgrenzen zu finden. Die Evaluierung dazu wird an einmal 8 Takten aus Bachs „Menuett in G-Dur“ und an 16 Takten aus Bachs „Marsch in D-Dur“ durchgeführt. Dies ist allgemein zu wenig, um die Leistungsfähigkeit eines Tonraums und eines Algorithmus bewerten zu können.

¹⁰ Vgl. Abschnitt 8.2.2.

¹¹ Noland und Sandler gehen ebenfalls von 12 Dur- und 12 Moll-Tonarten aus, damit von 24 Tonarten.

¹² Es existiert eine Datenbank mit der vollständigen Akkordbeschreibung aller Beatles-Alben. Diese wurde von Christopher Harte erstellt. Vgl. [57],[1].

¹³ Vgl. Abschnitt 3.6.

Hinzu kommt, dass die symbolische Beschreibung der Musikstücke, die für die Ergebnisbewertung notwendig ist, nicht eindeutig ist. In der Publikation „Polyphonic Audio Key Finding Using the Spiral Array CEG Algorithm“ [24] vergleicht Chew ihr Tonraummodell (und ihren Tonartalgorithmus CEG – *Center of Effect Generator*) mit der Leistungsfähigkeit von KK- und TY-Tonartprofilen. Es wird die Tonarterkennungsrate über die ersten 15 Sekunden von 28 Mozart-Symphonien ermittelt. Dabei schneiden alle Algorithmen nach 15 Sekunden ähnlich ab. Chew zeigt, dass ihr Algorithmus nach den ersten 3 Sekunden Analysezeit wesentlich besser ist als die anderen Algorithmen. Allerdings hat dieses Ergebnis keine Aussagekraft, weil nach den ersten 3 Sekunden höchstens der erste Akkord eines Musikstückes erklingen ist, welcher nicht immer die Tonika sein muss. Zudem weisen Mozart-Symphonien oft klare Strukturen auf, welche die Erkennung von Tonarten begünstigen. Ein weiterer Artikel „Audio Key Finding: Considerations in System Design and Case Studies on Chopin’s 24 Preludes“ [21] beschreibt ihren Tonartalgorithmus mit einigen Modifikationen und zeigt die Tonarterkennungsfähigkeit anhand 24 Präludien von Chopin. Ein Vergleich mit anderen Algorithmen findet nicht statt. Das Tonraummodell von Chew und der dazugehörige Tonartalgorithmus (CEG), der das tonale Zentrum beschreiben und erkennen soll, sind mathematisch sehr komplex. Elaine Chew knüpft an ihr Modell viele Regeln und Einschränkungen, deren prinzipielle Wirkungsweise z.T. nur punktuell gezeigt und belegt wird.

Die Tonarterkennung mit symbolischem Audio kann als Vorstufe zur Tonarterkennung mit realem Audio gesehen werden, welches die wichtigere und praxistaugliche Anwendung ist. Die Tonarterkennungsalgorithmen mit symbolischem Audio sind in der Regel nicht einfach auf die Erkennung mit realem Audio übertragbar, da hierbei vor der eigentlichen Analyse des Musikstücks alle Noten zuerst noch geschätzt werden müssen. Es folgt die Tonarterkennung mit realem Audio.

9.1.3 Tonarterkennung mit realem Audio

Bei der Tonarterkennung mit realem Audio wird ein aufgezeichnetes Musiksignal (PCM-Daten) analysiert. In Abbildung 9.5 ist dies als Blockschaltbild skizziert. Der Unterschied zur Tonarterkennung mit symbolischem Audio liegt vor allem im Frontend. Allerdings unterscheidet sich der Inhalt der Merkmalsvektoren durch die Bearbeitungsstufen im Frontend erheblich von den einfach zu bestimmenden Merkmalsvektoren bei der Tonarterkennung mit symbolischem Audio. Deshalb sind die Algorithmen bei der Analyse ebenfalls sehr unterschiedlich.

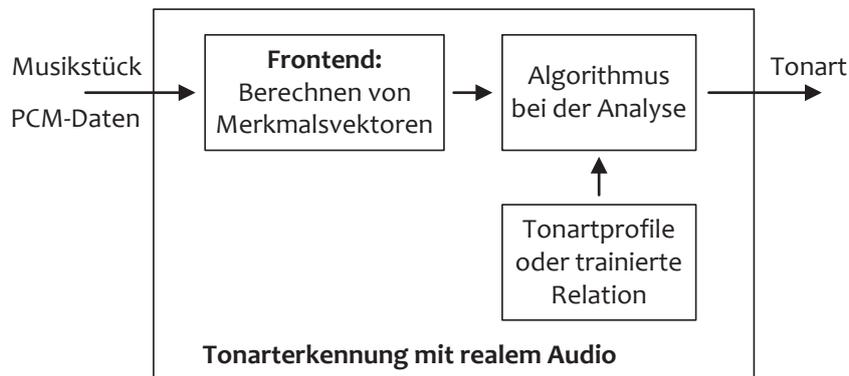


Abbildung 9.5 – Tonarterkennung mit realem Audio

Özgür İzmirlı verwendet eigene Tonartprofile, die er aus diatonisch gewichteten TY-Profilen¹⁴ und einem Referenztonset, das z.B. einzelne Klaviertöne beinhaltet, erstellt. In der Veröffentlichung „An Algorithm for Audio Key Finding“ [72] berechnet İzmirlı zu allen Chromavektoren (Merkmalsvektoren) der ersten 30 Sekunden des zu analysierenden Musikstücks einen Korrelationswert mit Hilfe seiner eigenen Tonartprofile. Aus den beiden wahrscheinlichsten Tonarten bildet er pro Chromavektor ein Konfidenzmaß und speichert dies zusammen mit der wahrscheinlichsten Tonart. Die größte Summe der Konfidenzmaße pro Tonart gibt das Ergebnis vor. Die Evaluierung des Algorithmus fand bei der MIREX2005¹⁵ statt. Hier schnitt der Algorithmus mit 89 % Erkennungsrate als bester ab. Anzumerken ist, dass die Teststücke bei der MIREX2005 ausschließlich synthetisierte Musikstücke waren, d.h. die Stücke wurden aus MIDI-Files künstlich erzeugt. Erkennungsalgorithmen sind in der Regel wesentlich besser bei der Verwendung von synthetisierten Musikstücken als bei realem Audio. Im Artikel „Audio Key Finding using Low-Dimensional Spaces“ [35] vergleicht İzmirlı seinen Tonarterkennungsalgorithmus mit einem Algorithmus, der vor der Konfidenzberechnung die Anzahl der Dimensionen des Merkmalsvektors reduziert. Dabei verwendet er die PCA (Principal Component Analysis¹⁶) und kommt zum Ergebnis, dass die Chromavektoren auf bis zu drei Dimensionen reduziert werden können, ohne dass sich die Tonarterkennungsrate merklich verschlechtert. İzmirlı hat damit gezeigt, dass die wesentlichen tonartbildenden Akkorde aus drei

¹⁴ TY-Profile besitzen wie KK-Profile für alle Tonigkeiten einen Wert ungleich Null. Diatonisch gewichtete Profile werden bei den tonleiterfremden Tonigkeiten zu Null gesetzt, behalten aber für die tonleitereigenen Tonigkeiten den ursprünglichen Wert.

¹⁵ Die MIREX (*Music Information Retrieval eXchange*) ist ein Wettbewerb zum Vergleich von Algorithmen. Sie findet seit 2005 jährlich als Begleitveranstaltung zur ISMIR (*International Symposium on Music Information Retrieval*) statt.

¹⁶ Engl.: Principal Component Analysis – Hauptkomponentenanalyse.

verschiedenen Tonigkeiten bestehen, was zu erwarten war, da die Dur- und Molldreiklänge die Grundlage für die Harmoniebildung darstellen. In seiner Arbeit „Localized Key Finding from Audio Using Non Negative Matrix Factorization for Segmentation“ [37] versucht er, ein Musikstück in mehrere Teile zu zerlegen, bevor für die einzelnen Teile jeweils die Tonart ermittelt wird. Es wird angenommen, dass ein Musikstück mehrere Tonarten enthalten kann. Daher ist es sinnvoll, die Bereiche unterschiedlicher Tonart vorher voneinander zu trennen. Der eigentliche Tonarterkennungsalgorithmus ist derselbe wie oben beschrieben. Die Zerlegung des Musikstückes wird mit NNMF¹⁷ erreicht. Dabei wird das Chromagramm (Matrix aus Chromavektoren) der zu analysierenden Musikpassage in ein Produkt zweier Matrizen zerlegt, die nicht negative Werte enthalten. Die erste Matrix enthält nach der Zerlegung Chromavektormuster, die linear unabhängig und deshalb möglichst unterschiedlich aussehen sollten. Die zweite Matrix gibt an, wie die Chromavektormuster zu mischen sind, damit das originale Chromagramm wieder hergestellt werden kann.¹⁸ In der zweiten Matrix sind dadurch eindeutig zeitliche Übergänge zu erkennen, nämlich genau dort, wo die Mischgewichte sich sprunghaft ändern. Dadurch wird eine Segmentierung des Musikstückes¹⁹ erreicht. İzmirli verwendet zur Evaluierung Musikstücke aus den Bereichen Pop und Klassik. Die Ergebnisse zeigen keine Verbesserung zu den bisherigen Veröffentlichungen, da hier hauptsächlich die Segmentierung mittels NNMF getestet und kein neuer Tonarterkennungsalgorithmus vorgestellt wurde. Lee und Seung haben 1999 NNMF erstmals für die Bildsegmentierung vorgeschlagen (vgl. [92]).

Steffen Pauws nutzt in seinem Artikel „Musical Key Extraction from Audio“ [122] spezielle Chromavektoren, die er als “harmonically compressed”²⁰ bezeichnet. Hierbei werden im Spektrum die ersten 15 Obertöne zu jedem Grundton gewichtet hinzuaddiert, bevor die Chromavektoren erstellt werden. Die Tonarterkennung wird durch einfache Korrelation mit KK-Tonartprofilen durchgeführt. Die Ergebnisse bei den angegebenen Teststücken liegen bei ca. 70 % Erkennungsrate und damit unterhalb der besten Ergebnisse anderer Ansätze.

¹⁷ Engl: Non negative matrix factorization – Nicht negative Matrix Faktorisierung.

¹⁸ Der NNMF-Algorithmus und die Initialisierung dieses Algorithmus beeinflussen das Ergebnis der Faktorisierung sehr stark. Dadurch ist auf diese Punkte bei weiteren Untersuchungen Wert zu legen.

¹⁹ Das Musikstück wird zwar segmentiert. Es kann jedoch sein, dass eine Segmentgrenze an einer Stelle gesetzt wird, an der sich z.B. die Instrumentierung und nicht die Tonart ändert.

²⁰ Engl.: harmonically compressed – harmonisch komprimiert.

Emilia Gómez verwendet sogenannte „Harmonic Pitch Class Profiles“ (HPCP) in ihrer Veröffentlichung „Key Estimation from Polyphonic Audio“ [52] zur Tonarterkennung. HPCP sind erweiterte Chromavektoren, welche eine dreifache Halbtonauflösung haben. Aus den TY-Profilen berechnet sie eigene Tonartprofile, die ähnlich wie bei Steffen Pauws die Obertonstruktur eines Klanges berücksichtigen. Außerdem werden die Töne der drei Grundakkorde (Tonika, Subdominante und Dominante) stärker gewichtet. Die Evaluierung des Algorithmus erfolgte bei der MIREX2005. Hier erreichte der Algorithmus eine Erkennungsrate von 85 %.

Geoffroy Peeters nutzt – ebenso wie Noland und Sandler – HMM zur Erkennung von Tonarten. In der Publikation „Musical Key Estimation of Audio Signals based on Hidden Markov Modeling of Chroma Vectors“ [123] wird jeweils ein HMM für Dur und ein HMM für Moll erstellt, die auf Chromavektor-Basis arbeiten. Durch Rotation der Vektoren und Matrizen werden die HMM für alle Tonigkeiten (Tonarten) erzeugt. Das Training ist EM-basiert und Peeters verwendet drei, sechs oder zwölf Zustände (Chromavektoren als Mittelwerte plus Kovarianzmatrizen), die nicht näher definiert sind und deren Inhalte durch das Training bestimmt werden.²¹ Peeters untersucht, ob eine oder drei Gaußmischungen für die Modellierung des Wahrscheinlichkeitsschätzers besser geeignet sind. Die Ergebnisse liegen bei allen Kombinationen zwischen 81 % und 85 %. Es zeigt sich, dass die Zustände nach dem EM-Training fast alle den gleichen Inhalt haben. Dadurch geht die wichtigste Eigenschaft des HMM, die zeitliche Relation, verloren, da ein Zustandswechsel keine Veränderung hervorruft, denn das HMM bleibt damit immer im gleichen Zustand. Das HMM fällt in einen einfachen Korrelationsalgorithmus zusammen. Die Ursache für dieses Problem ist das nicht überwachte EM-Training.

Alexander Lerch beschreibt in seinem Artikel „Ein Ansatz zur automatischen Erkennung der Tonart in Musikdateien“ [100] einen einfachen und effektiven Tonarterkennungsalgorithmus. Er berechnet von allen zu analysierenden Chromavektoren (Merkmalsvektoren) eines Musikstücks den Mittelwert und bestimmt anschließend die euklidische Distanz zu den 24 KK-Tonartprofilen. Die geringste Distanz gibt die Tonart an. Die Erkennungsrate für 210 Popstücke wird mit 77 % angegeben, wobei hier nur die korrekten Treffer aufgelistet sind. Seit der MIREX2005 gibt es den sogenannten MIREX-Score²², der auch

²¹ Das Training ist damit nicht überwacht (engl.: „unsupervised learning“) und kann je nach Initialisierung und Auswahl der Teststücke jedesmal sehr unterschiedlich ausfallen. Ebenso ist es möglich, dass nach dem Training eine perfekte Korrelation zwischen den Zuständen existiert. Alle Zustände sind identisch.

²² Vgl. Kapitel 11.

Teiltreffer vorsieht, wenn Nachbartonarten (z.B. im Quintabstand) erkannt werden. Das Ergebnis des Algorithmus ist nach der Bewertung mit dem MIREX-Score höher als 77 % anzusetzen.

Zhu, Kankanhalli und Gao verfolgen in ihrer Veröffentlichung „Music Key Detection for Musical Audio“ [168] einen zweistufigen (hierarchischen) Ansatz. Im ersten Schritt wird die diatonische Tonleiter des Musikstücks ermittelt. Anschließend wird innerhalb der diatonischen Tonleiter bestimmt, ob es sich um ein Dur- oder Moll-Stück handelt. Die Bestimmung der diatonischen Tonleiter bezeichnen die Autoren mit dem Begriff Konsonanzfilterung. Dabei werden pro Analysezeitfenster mit der stärksten Spektrallinie beginnend so lange alle weiteren Spektrallinien zum aktuellen Chromagramm addiert, bis eine Spektrallinie diatonisch nicht mehr dazu passt, d.h. nach jeder Addition wird untersucht, ob alle Tonigkeiten mit einer Energie ungleich Null zu einer einzigen diatonischen Tonleiter gehören. Wenn dies nicht der Fall ist, so wird die Konsonanzfilterung für das aktuelle Zeitfenster und den damit verknüpften Chromavektor beendet. Es werden alle Chromavektoren der zu analysierenden Musikpassage addiert. Anschließend wird die diatonische Tonleiter ausgewählt, welche die größte Energie besitzt. Innerhalb der diatonischen Tonleiter gibt wiederum diejenige Quinte mit der größten Energie die Tonart vor, wobei allerdings nur zwischen Dur und Moll unterschieden wird. Die Evaluierung erfolgt mit 60 nicht näher benannten Popstücken und 12 klassischen Stücken aus Vivaldis „Vier Jahreszeiten“. Bei Pop erreicht der Algorithmus 90 % und bei Klassik 83 %. Es wird kein Vergleich mit anderen Tonarterkennungsalgorithmen durchgeführt.

Die hier dargestellten Ergebnisse und die Leistungsfähigkeit der verschiedenen Ansätze sind nur sehr schwierig zu bewerten, da ein echter Vergleich zwischen den Algorithmen fehlt. Bei der MIREX2005 wurden bisher ein einziges Mal unterschiedliche Algorithmen mit den gleichen zu analysierenden Musikstücken getestet. Allerdings wurde hier nur synthetisiertes Audio und kein reales Audio verwendet. In den Veröffentlichungen werden die zu analysierenden Musikstücke meistens nicht benannt oder die Anzahl der genannten Stücke ist für eine statistisch gesicherte Aussage zu gering. Da die Anzahl an Musikstücken unüberschaubar groß ist, finden sich immer genügend einzelne Stücke, die für den zu bewertenden Algorithmus einfach zu analysieren sind. Ein Vergleich zu einigen hier dargestellten Algorithmen und dem in dieser Arbeit neuen Algorithmus ist in Kapitel 11 zu finden.

9.2 Akkorderkennung

Dieser Abschnitt stellt den Stand der Technik zur Akkorderkennung dar. Alle hier vorgestellten Algorithmen arbeiten mit realem Audio.

Sheh und Ellis verwenden als eine der ersten HMM im Bereich der Akkorderkennung. In ihrem Artikel „Chord Segmentation and Recognition using EM-Trained Hidden Markov Models“ [148] besteht die Hauptaufgabe zwar in der Ausrichtung von Akkordgrenzen, jedoch wird die Akkorderkennung ebenfalls betrachtet. Für die Ausrichtung der Akkordgrenzen sind die Akkorde zwar bereits bekannt, es ist aber unbekannt, an welcher Stelle sich diese befinden. Dies ist ein vollkommen anderer Ansatz als das Erkennen von Akkordgrenzen ohne jegliches Vorwissen. Für die genannte Aufgabe verwenden Sheh und Ellis den EM-Algorithmus. Das Training erfolgt nicht überwacht und die Initialisierung der Matrizen ist zufällig. Sie schlagen auch als erste vor, dass die Mittelwerte der Wahrscheinlichkeitsschätzer bei Verwendung von Gaußmischverteilungen bei der gleichen Akkordklasse (z.B. alle Dur-Akkorde oder alle Moll-Sept-Akkorde) für alle zwölf Tonigkeiten gemeinsam verwendet werden können. Dadurch erhöht sich die Anzahl der Trainingsmerkmalsvektoren um ein Vielfaches. Wenn z.B. der Mittelwertsvektor vom Zustand (Akkord) C-Dur um zwei Tonigkeiten nach oben rotiert wird, so kann dieser als Mittelwertsvektor für den Zustand (Akkord) D-Dur verwendet werden. Somit können alle Akkorde der gleichen Akkordklasse gemeinsam einen trainierten Mittelwertsvektor bilden, der für alle zwölf Tonigkeiten verwendet wird. Zur Evaluierung benutzen Sheh und Ellis jeweils 18 Beatles-Lieder für das Training und zwei andere Beatles-Lieder für die Analyse. Sie vergleichen ihre Ergebnisse mit anderen Merkmalsvektoren, wobei jedoch die Chromavektoren mit der beschriebenen Rotation bei beiden Aufgaben die besten Ergebnisse liefern. Bei der Ausrichtung der Akkordgrenzen erkennt der Algorithmus 75 % aller Grenzen und bei der Akkorderkennung werden 22 % Erkennungsrate erzielt. Zwei Teststücke sind allerdings für eine Evaluierung zu wenig. Damit sind die Ergebnisse unsicher.

Shenoy und Wang verfolgen in ihrem Artikel „Key, Chord and Rhythm Tracking of Popular Music Recordings“ [151] einen iterativen Ansatz zur Akkorderkennung. Dabei nutzen sie bekannte Algorithmen aus dem Bereich der Tonart- und der Rhythmuserkennung. Mit Hilfe des Rhythmus' lassen sie nur Akkordwechsel auf den Zählzeiten eins und drei eines Vierviertel-Taktes zu; mit Hilfe der Tonart werden tonartfremde Akkorde gelöscht bzw. unsichere Akkorde durch Nachbarakkorde interpoliert. Dazu ist im Vorfeld eine Akkordschätzung notwendig, die aus einer anderen Veröffentlichung von Shenoy verwendet

wird. Bei 30 untersuchten Teststücken aus dem Bereich Pop wird eine Akkorderkennungsrate von 48 % bei der ersten Akkordschätzung erreicht. Nach der Anwendung des Algorithmus mit Rhythmus- und Tonarterkennung verbessert sich die Erkennungsrate auf 79 %. Die Evaluierung zeigt mehrere Schwachpunkte, und die Ergebnisse sind daher kritisch zu betrachten: Es werden ausschließlich Stücke mit Vierviertel-Takt analysiert, was nicht repräsentativ ist. Zudem wird vorausgesetzt, dass sich die Tonart eines Stückes nicht ändert. Das trifft nur auf einen Teil der vielen Pop-Musikstücke zu. Der Ansatz kann daher bei Stücken, die ihre Tonart wechseln, nicht funktionieren, es sei denn, der Tonarterkennungsalgorithmus ist in der Lage, einen Tonartwechsel zu erkennen. Der Stand der Technik zur Tonarterkennung gibt dies allerdings nicht her. Des Weiteren ist es sehr wohl möglich, dass ein Akkord gespielt wird, der nicht Bestandteil einer Tonart ist. Dies trifft auf Doppeldominanten und Zwischendominanten zu, die durch den hier vorgestellten Algorithmus entfernt würden. Der vorgestellte Ansatz mit seinen Einschränkungen ist sehr speziell und funktioniert daher nur für eine sehr eingegrenzte Auswahl an Musikstücken.

Harte und Sandler benutzen in ihrer Publikation „Automatic Chord Identification Using a Quantised Chromagram“ [157] perfekte Bitmuster-Chromavektoren zur Akkordanalyse, d.h. der „trainierte“ Vektor für z.B. C-Dur ist [1 0 0 0 1 0 0 1 0 0 0 0] und trägt damit für die akkordeigenen Tonigkeiten den Wert „1“ und für die restlichen den Wert „0“. Es werden je 12 Akkordmuster für Dur-Akkorde, Moll-Akkorde, verminderte Akkorde und übermäßige Akkorde angelegt. Das größte Skalarprodukt aus dem zu analysierenden Merkmalsvektor und den Chromavektormustern gibt den erkannten Akkord vor. Zur Evaluierung werden die beiden Beatles-Alben *Please Please Me* (CD1) und *Beatles for Sale* (CD2) verwendet. Für CD1 erreicht der Algorithmus eine Erkennungsrate von 53,9 % und für CD2 70,8 %. Die Auswertung ergibt, dass die meisten Fehler durch harmonisch ähnliche Akkorde passieren, so wird z.B. *h*-Moll anstelle von *D*-Dur erkannt.

Sailer und Rosenbauer wählen in ihrem Artikel „A Bottom-Up Approach to Chord Detection“ [137] einen Ansatz, der für jeden zu analysierenden Merkmalsvektor solange Akkordkandidaten wegstreicht, die unwahrscheinlich sind, bis ein einzelner Akkord übrigbleibt. Dafür werden zuerst aus dem Spektrum die sichersten Töne extrahiert und den möglichen Akkordkandidaten zugeordnet. Mit der Tonart, die mit KK-Profilen geschätzt wird, werden wie bei Shenoy und Wang weiterhin viele Akkorde ausgeschlossen. Wenn zwei Akkorde gleichwahrscheinlich sind, so wird derjenige Akkord als richtig angenommen, der zu den Hauptfunktionen gehört. Der Algorithmus liefert bei sechs Pop-Stücken eine Erkennungsrate von ca. 30 %. Dies liegt unter den Ergebnissen der anderen Ansätze.

Es wird kein Vergleich zu anderen Algorithmen hergestellt. Für eine gesicherte Aussage zur Leistungsfähigkeit des Algorithmus sind sechs Stücke zu wenig. Hinzu kommt, dass der hier vorgestellte Ansatz Schwachpunkte hat: Erstens werden nur tonleitereigene Akkorde zugelassen, so dass hier die gleiche gegensätzliche Begründung wie zum Ansatz von Shenoy und Wang gilt. Zweitens ist die Tonarterkennung nicht evaluiert worden, so dass dann, wenn eine falsche Tonart erkannt worden ist, die hier vorgestellte und darauf aufbauende Akkorderkennung keine guten Ergebnisse erzielen kann.

Bello und Pickens verwenden in ihrem Ansatz „A Robust Mid-level Representation for Harmonic Content in Music Signals“ [12] HMM für die Akkorderkennung. Die Parameter A , b und π werden mit dem EM-Algorithmus trainiert, wobei die Initialisierungen im Gegensatz zum Algorithmus von Sheh und Ellis nicht zufällig sind. Für die Werte der Zustandsübergangsmatrix A wird zur Initialisierung der doppelt verschachtelte Quintenzirkel benutzt. Akkorde, die geometrisch nahe beieinander liegen, werden damit wahrscheinlicher als Folgeakkord angenommen als Akkorde, die weiter entfernt liegen. Als Mittelwertsvektoren von b^{23} werden ähnlich wie bei Harte und Sandler, die ohne HMM und damit ohne zeitliche Relation arbeiten, perfekte Chromavektormuster gewählt. Die Kovarianzmatrizen werden jeweils auf drei verschiedene Arten initialisiert. Bei der ersten sind es reine Diagonalmatrizen²⁴ und alle Werte auf den Diagonalen sind gleich groß. Bei der zweiten sind es ebenfalls reine Diagonalmatrizen, aber die Varianzen der Tonigkeiten, die nicht zum Akkordmuster gehören, werden sehr klein gewählt. Die dritte Art ist ähnlich der zweiten. Es werden noch zusätzlich die Kovarianzen der drei Tonigkeiten des Akkordmusters mit Werten ungleich Null belegt. Die zu analysierenden Akkordklassen werden auf reine Dur- und Moll-Akkorde beschränkt. Bei der Evaluierung werden verschiedene Kombinationen der HMM-Parameter mit dem EM-Algorithmus trainiert. Gleichzeitig wird der Rhythmus des zu analysierenden Stückes geschätzt und zur zeitlichen Segmentierung verwendet. Es zeigt sich, dass bei Anwendung des Rhythmus' und des Trainings von A und π die besten Ergebnisse erzielt werden. Bello und Pickens nehmen als Teststücke wie Harte und Sandler die beiden Beatles-Alben *Please Please Me* (CD1) und *Beatles for Sale* (CD2). Die Ergebnisse beider Untersuchungen sind damit vergleichbar; Bello und Pickens erhalten bessere Ergebnisse: Für CD1 erreichen sie als bestes Ergebnis 68,6 % und für CD2 81,6 %. Der Mittelwert beträgt 75,0 %. Es ist anzumerken, dass das Training der Kovarianzmatrizen immer problematisch ist, da die Kovarianzen wesentlich mehr Trainingsstücke benötigen als die Varianzen. Die Werte der Diagona-

²³ Für den Wahrscheinlichkeitsschätzer b werden einzelne zu trainierende Gaußverteilungen benutzt.

²⁴ Eine Diagonalmatrix ist eine Matrix, die nur auf der Diagonalen Werte ungleich Null enthält.

len sind deshalb immer besser trainiert als die übrigen Werte der Diagonalmatrix. Die meisten Algorithmen verwenden deshalb für die Analyse nur die Diagonalelemente.

Papadopoulos und Peeters evaluieren in ihrer Veröffentlichung „Large-Scale Study of Chord Estimation Algorithms Based on Chroma Representation and HMM“ [121] ebenfalls die Parameter eines HMM. Zur Analyse verwenden sie 110 Beatles-Lieder mit einer Erkennungsrate von 70,96 %. Dies wird bei der Verwendung von Krumhansl-Profilen für die Übergangswahrscheinlichkeiten und einfacher Korrelation der zu analysierenden Merkmalsvektoren mit Akkordmustern zur Wahrscheinlichkeitsschätzung der Zustände erreicht. Es wird kein Vergleich zu anderen Algorithmen dargestellt.

Kyogu Lee benutzt zur Akkorderkennung in seinem Artikel „A System for Automatic Chord Transcription from Audio Using Genre-Specific HMM“ [95] mehrere HMM, die für unterschiedliche Genres trainiert werden. Das Training wird mit über 4000 synthetisierten MIDI-Stücken durchgeführt. Als Genres gibt Lee *Keyboard*, *Chamber*, *Orchestral*, *Rock*, *Jazz* und *Blues* an. Der 6D-Tonraum von Harte, Sandler und Gasser²⁵ wird für die Berechnung der Merkmalsvektoren verwendet. Die Teststücke sind sämtliche Lieder der beiden Beatles-Alben *Please Please Me* (CD1) und *Beatles for Sale* (CD2). Die Evaluierung zeigt, dass das Genre-spezifische HMM für *Rock* die besten Ergebnisse mit einer Erkennungsrate von 60,0 % für CD1 und 84,3 % für CD2 bei einem Mittelwert von 72,2 % erzielt. Die Ergebnisse sind denen von Bello und Pickens ähnlich. Es ist anzumerken, dass das Training zwar mit vielen Musikstücken durchgeführt wird, jedoch ist für ein besseres Training reales Audio zu verwenden. Außerdem findet kein Vergleich zu Chromavektoren als Merkmalsvektoren statt. In einer anderen Veröffentlichung „A Unified System for Chord Transcription and Key Extraction using Hidden Markov Models“ [94] verwendet Kyogu Lee zusammen mit Malcolm Slaney tonartbezogene HMM und verbessert die Erkennungsrate für CD1 auf 64,0 % und für CD2 auf 84,7 %. Der Mittelwert beträgt hier 74,35 %.

Kouhei Sumi und Kollegen wählen in ihrem Artikel „Automatic Chord Recognition Based on Probabilistic Integration of Chord Transition and Bass Pitch Estimation“ [95] einen Ansatz, der Rhythmus- und Basslinienerkennung mit Akkorderkennung verbindet. Während der Analyse werden mit trainierten Gaußmischverteilungen Akkorde geschätzt und eine Art Übergangswahrscheinlichkeit verwendet. Dies entspricht einem HMM, die Autoren nennen ihren Ansatz aber nicht HMM-basiert. Die Evaluierung wird mit 150

²⁵ Siehe Abschnitt 3.4 und Abschnitt 3.5.

Beatles-Liedern und mit den beiden Alben CD1 und CD2 durchgeführt. Für die 150 Lieder wird eine Erkennungsrate von 73,4 % erzielt und für die beiden einzelnen Alben eine Erkennungsrate von im Mittel 77,5 %. Die Ergebnisse sind damit etwas besser als die Ergebnisse der anderen vorgestellten Arbeiten.

Die hier präsentierten Ergebnisse zur Akkorderkennung für die beiden Beatles-Alben CD1 und CD2 sowie für die größere Auswahl von 150 Beatles-Liedern werden als Vergleich für die Evaluierung des in dieser Arbeit neuen Algorithmus' zur Akkorderkennung herangezogen. Die Ergebnisse sind in Kapitel 12 zu finden.

Kapitel 10

Symmetriemodellparameter bei realem Audio

Dieses Kapitel beschreibt für reales Audio die Ausgabeparameter des Symmetriemodells, das in Kapitel 4 als neues Tonraummodell definiert und eingeführt worden ist. Das Verstehen der Ausgabeparameter dient dazu, Ideen und Algorithmen umsetzen zu können, die auf diesen aufbauen. In Abschnitt 10.1 wird untersucht und diskutiert, wie ein Einzelton, ein Intervall, ein Akkord und eine Kadenz im Symmetriemodell repräsentiert werden. Anschließend wird in Abschnitt 10.2 die geometrische Lage von ganzen Musikstücken betrachtet. Dies geschieht mit Hilfe von Histogrammen und Verteilungen in den Tonräumen des Symmetriemodells.

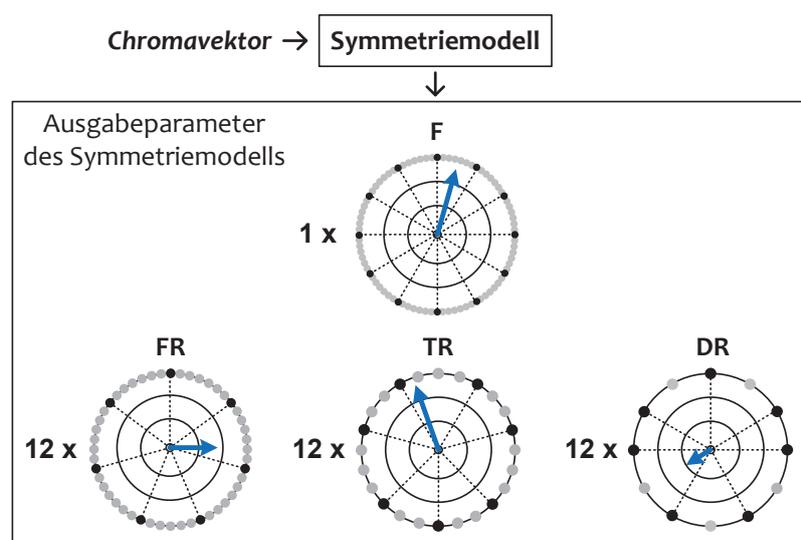


Abbildung 10.1 – Die 37 verwendeten Symmetriemodell-Kreise

Die Eigenschaften des Symmetriemodells werden an dieser Stelle kurz zusammenfassend erläutert: Das Symmetriemodell besteht aus mehreren Kreisen. Die hier verwendeten Kreise werden in Abbildung 10.1 dargestellt. Zu sehen sind der tonartübergreifende Quintkreis (Quintenzirkel) F , zwölf tonartbezogene Quintkreise FR , zwölf tonartbezogene Terzkreise TR und zwölf tonartbezogene diatonische Kreise DR . Es gibt für jede diatonische Tonleiter drei tonartbezogene Kreise, jeweils einen FR , einen TR und einen DR . Zur Auswahl des Kreises einer diatonischen Tonleiter wird der Einfachheit halber an den jeweiligen Kreis die ionische Tonart angehängt: So wird z.B. der TR der diatonischen Tonleiter, die C-Ionisch, D-Dorisch usw. beinhaltet, mit $TR(C)$ benannt. Abbildung 10.2 zeigt als Beispiel einen Chromavektor, der das harmonische Ereignis a -Moll beinhaltet. Dieses wird in den Kreisen $FR(C)$, $TR(C)$ und $DR(C)$ dargestellt. Die Eigenschaften der Kreise des Symmetriemodells sind die folgenden:

1. Es gibt einen Symmetrieton für jede diatonische Skala. Dieser wird in den Kreisen immer oben angeordnet. Die senkrechte Achse durch diesen Ton ist die Symmetrieachse, um die sich die musikalischen Ereignisse symmetrisch anordnen.
2. In jedem Kreis wird zu einem Ereignis (Chromavektor) aus den Vektoren der Einzeltonigkeiten ein Summenvektor gebildet, der bestimmte musikalische Eigenschaften widerspiegelt.
3. Der Winkel des Summenvektors drückt die Art eines harmonischen oder tonalen Ereignisses aus. Der Winkel 0° zeigt nach oben und wächst im Uhrzeigersinn positiv.
4. Die Länge des Summenvektors gibt an, wie stark das musikalische Ereignis ausgeprägt ist.

Das Beispiel in Abbildung 10.2 zeigt, dass der Summenvektor für einen a -Moll-Dreiklang im $TR(C)$ am längsten ist. Dies ergibt sich, da die Tonigkeiten des Dreiklanges im $TR(C)$ direkt nebeneinander liegen. Der TR kann allgemein Terzen, Dreiklänge und Vierklänge in Terzschichtung erkennen. Dabei ist die Länge des Vektors im Verhältnis zu den anderen Kreisen zu betrachten. Im $FR(C)$ ist der Summenvektor für das gezeigte Beispiel wesentlich kürzer. Die Summe der Tonigkeitsvektoren a und e ist zwar relativ lang, jedoch liegt die Terz c direkt gegenüber und verkürzt den Summenvektor. Der FR kann wegen seiner Quintschichtung vor allem Sus-Akkorde erkennen und trennt Dur- und Moll-Akkorde, die in der linken und rechten Hälfte des Kreises liegen. Die Symmetrieachse ist hierbei die Teilungsachse. Der Summenvektor im $DR(C)$ ist für dieses Beispiel sehr kurz, da sich die Vektoren der Tonigkeiten eines Drei- oder Vierklanges nahezu gegenseitig aufheben.

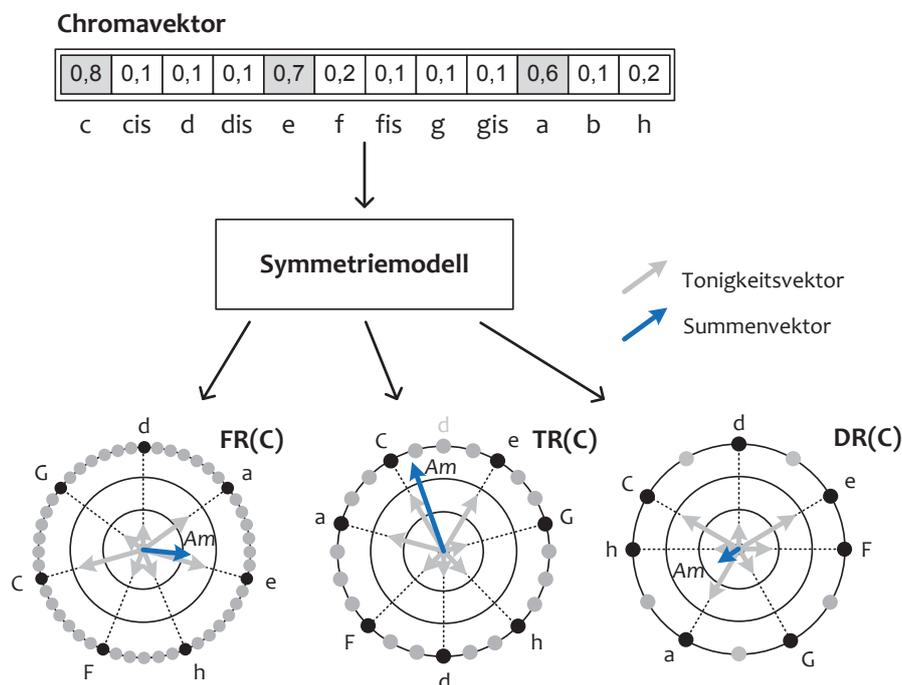


Abbildung 10.2 – Repräsentation des Ereignisses *a*-Moll in den Kreisen von C-Dur

Die Länge des Summenvektors im *DR* ist dazu geeignet, im Verhältnis zu den anderen beiden Summenvektorlängen Einzeltöne und Ton-Cluster zu erkennen.

10.1 Repräsentation von Tönen, Intervallen und Akkorden

In diesem Abschnitt wird gezeigt, wie sich die Ausgabeparameter des Symmetriemodells anhand von Orgelklängen für die Ereignisse Einzelton, Intervall, Akkord und Kadenz im zeitlichen Verlauf verhalten. Des Weiteren zeigen Punktwolken zu den nachfolgenden Beispielen in den Kreisen *FR*, *TR* und *DR* die geometrische Lage der Ereignisse.

10.1.1 Einzelton

Abbildung 10.3 zeigt einen Einzelton, den Ton *c*. Im Diagramm (a) ist das Zeitsignal zu sehen und damit auch die Zeitbereiche des Tonanschlags („Attack“), des Klingens („Sustain“) und des Tonloslassens („Release“).

Diagramm (c) zeigt für den Ton aus Diagramm (a) die Betragsverläufe der Summenvektoren in den Kreisen *FR(C)*, *TR(C)* und *DR(C)*. Für einen Einzelton sind die Beträge während der Sustain-Phase in den drei Kreisen ungefähr gleich groß, da nur eine Tonigkeit eine große Amplitude besitzt, nämlich die erklingende Tonigkeit. Die Summierung

mit den restlichen Tonigkeitsvektoren, die zwar in jedem Kreis an anderen Stellen sind und dadurch verschiedene Winkel, jedoch keine große Amplitude haben, bewirkt keine wesentliche Veränderung des Vektors.

Solange das Zeitsignal rauscht und eine sehr kleine Amplitude hat, ist der Betrag des Summenvektors in jedem Kreis klein und die Winkel variieren zufällig. Dies ist im Diagramm (b) in Abbildung 10.3 zu sehen. Die Amplitude eines Summenvektors gibt an, wie zuverlässig dessen dazugehöriger Winkel und damit das harmonische Ereignis ist. Wenn alle drei Vektoren der Kreise *FR*, *TR* und *DR* während der Sustain-Phase gleich lang sind, dann lässt sich in jedem Kreis mit dem Winkel direkt die erklingende Tonigkeit ermitteln. Das Diagramm (d) in Abbildung 10.3 zeigt in den drei Kreisen *FR(C)*, *TR(C)* und *DR(C)* Punktwolken des Ereignisses. Dabei entspricht jeder Punkt einem Summenvektor und damit einem Zeitfenster (128 ms) des Signals. Der *FR(C)* gibt an, dass der Einzelton *c* in C-Dur einem Dur-Ereignis zugeordnet wird. Der *TR(C)* zeigt, dass der Ton *c* zum tonischen Bereich gehört.

10.1.2 Intervall

Die Abbildungen 10.4 und 10.5 zeigen die Ereignisse „große Terz“ (*c-e*) und „reine Quinte“ (*c-g*). Die Quinte kann auch als C-Dur-Akkord ohne Terz und damit als C-sus-Akkord interpretiert werden.

Für die Terz *c-e* in Abbildung 10.4 ist der Betrag des Summenvektors im *TR(C)* am längsten und der Betrag des Summenvektors im *DR(C)* am zweitlängsten. Die Terz wird im Symmetriemodell keinem Tongeschlecht (Dur/Moll) zugeordnet. Dies ist an der Punktwolke, die sich direkt auf der Symmetrieachse im *FR(C)* (Abbildung 10.4, Diagramm (d)) befindet, zu sehen. Die Zuordnung ist musikalisch richtig, da die Terz sowohl Teil eines C-Dur-Akkordes als auch Teil eines *a*-Moll-Akkordes sein kann. Die Punktwolke im *TR(C)* zeigt, dass die Terz *c-e* in dieser Tonart zum tonischen Bereich gehört.

Die reine Quinte in Abbildung 10.5 erzeugt den längsten Summenvektor im *FR(C)* und den zweitlängsten im *TR(C)*. Weil der *TR* durch nebeneinander liegende Terzen aufgebaut ist, befinden sich harmonische Ereignisse, die terzverwandt sind, nahe beieinander. Dadurch sind die Punktwolken der Terz *c-e* und der Quinte *c-g* im *TR* nicht weit voneinander entfernt (vgl. Abbildung 10.4 und Abbildung 10.5, jeweils Diagramm (d)). Der Winkel des Summenvektors ist im *FR(C)* negativ (vgl. Abbildung 10.5). Die Quinte, C-sus-Akkord, wird damit einem Dur-Akkord zugeordnet. Da Sus-Akkorde kein eindeutiges Tongeschlecht haben, werden sie in Abhängigkeit vom musikalischen Kontext entsprechend wahrgenommen und eingeordnet. Der Summenvektor im *DR* ist kurz, weil die Tonigkeiten *c* und *g* im *DR(C)* fast gegenüberliegen (vgl. Abbildung 10.5, Diagramm (d)).

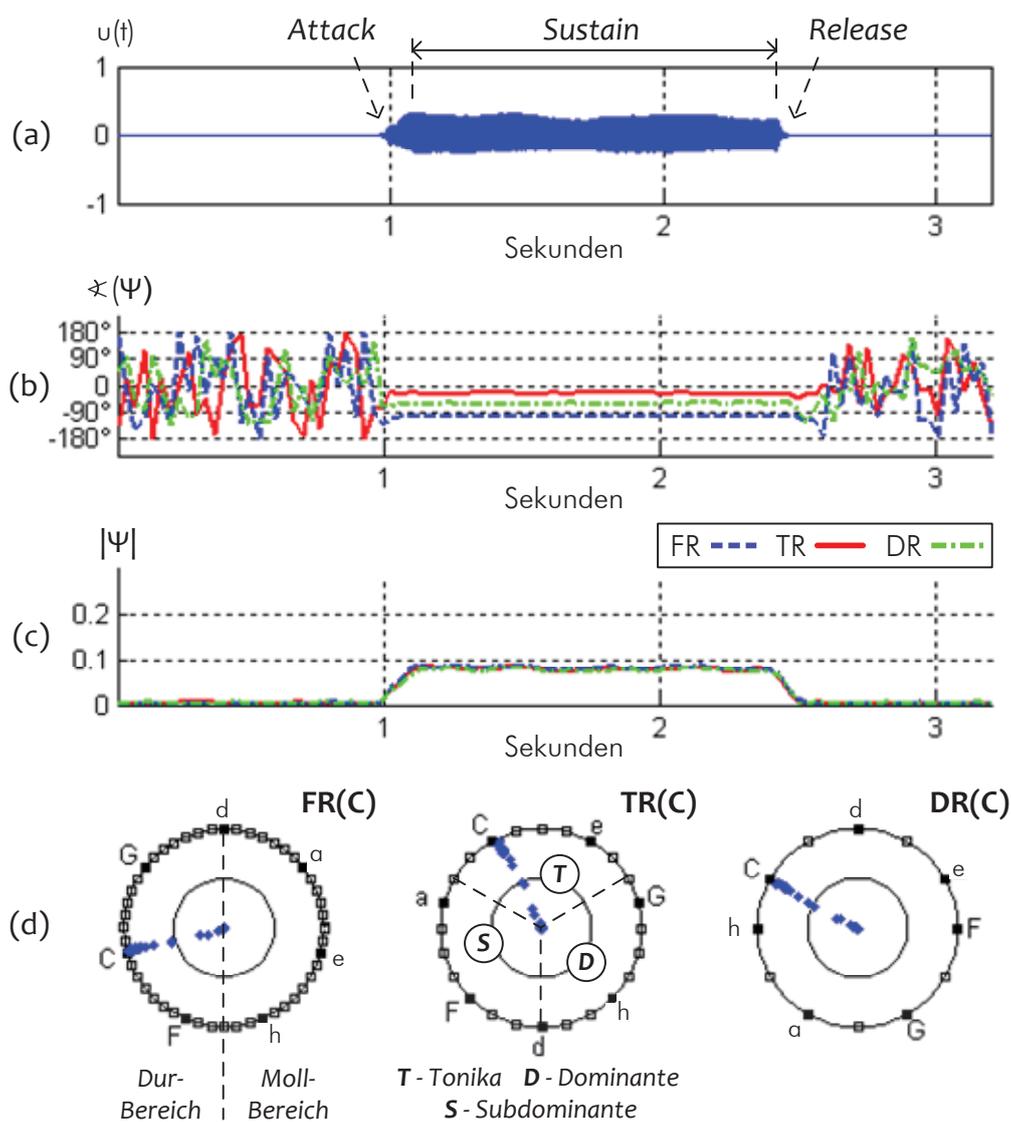


Abbildung 10.3 – Einzelton *c*, Orgel-Klang; (a) Signal im zeitlichen Verlauf; (b) Winkelverlauf in FR, TR und DR; (c) Betragsverlauf in FR, TR und DR; (d) Punktwolken in FR, TR und DR;

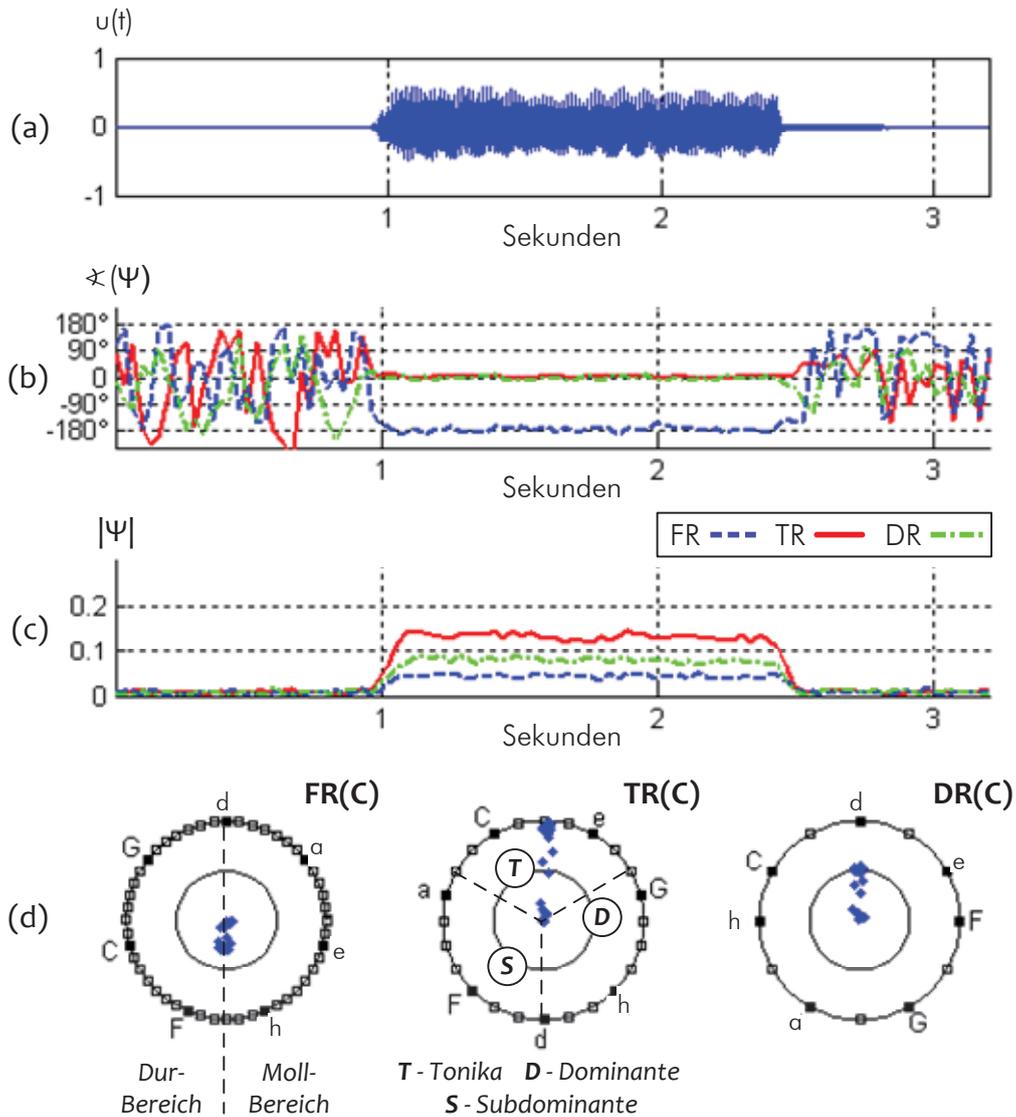


Abbildung 10.4 – Intervall Terz $c-e$, Orgel-Klang; (a) Signal im zeitlichen Verlauf; (b) Winkelverlauf in FR, TR und DR; (c) Betragsverlauf in FR, TR und DR; (d) Punktwolken in FR, TR und DR

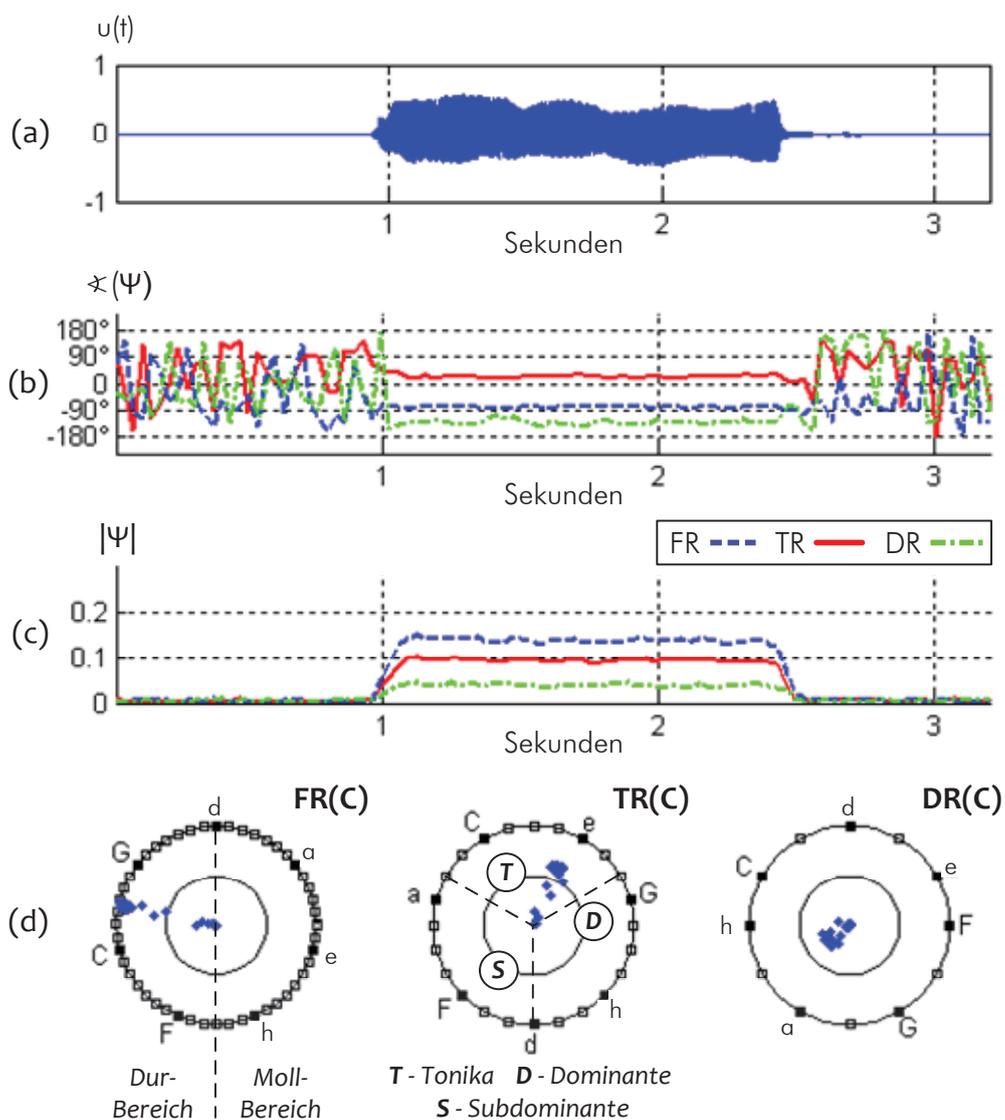


Abbildung 10.5 – Intervall Quinte *c-g*, Orgel-Klang; (a) Signal im zeitlichen Verlauf; (b) Winkelverlauf in FR, TR und DR; (c) Betragsverlauf in FR, TR und DR; (d) Punktwolken in FR, TR und DR

10.1.3 Akkord

In Abbildung 10.6 werden die Ausgabeparameter des Akkords *a*-Moll gezeigt. Da es sich hierbei um einen Dreiklang handelt, ist der Summenvektor im $TR(C)$ am längsten, der Summenvektor im $FR(C)$ am zweitlängsten. Im Diagramm (b) und Diagramm (d) ist zu sehen, dass der Winkel des Summenvektors im $FR(C)$ positiv ist. Der Akkord *a*-Moll wird daher dem Tongeschlecht Moll richtig zugeordnet. Der Akkord gehört zum tonischen Bereich, was der $TR(C)$ anzeigt. Sowohl Sus-Akkorde als auch Dreiklänge ergeben im DR stets kurze Summenvektoren. Dies ist für dieses Beispiel im Diagramm (d) zu sehen.

10.1.4 Grundkadenz

Die Grundkadenz in der Tonart C-Dur besteht aus einer Folge der Akkorde C-Dur, F-Dur, G-Dur und wiederum C-Dur. Abbildung 10.7 zeigt das Zeitsignal der Kadenz in der Tonart C-Dur sowie die Ausgabeparameter des Symmetriemodells für die gespielte Kadenz. Die Winkel sind während der Sustain-Phasen im $TR(C)$ und $FR(C)$ sehr stabil (vgl. Diagramm (b)). Durch eine Überbetonung der Tonigkeit *a* in der zweiten Phase des F-Dur-Akkordes verändern sich zwar die Beträge der Summenvektoren (vgl. Diagramm (c)), die Winkel bleiben jedoch auch für dieses Teilereignis im $TR(C)$ und $FR(C)$ konstant. Das Diagramm (d) zeigt, dass die gesamte Kadenz im Dur-Bereich liegt. Zugleich werden die funktionellen Bereiche Tonika, Dominante und Subdominante im $TR(C)$ durch die Punktwolken entsprechend ausgefüllt. Die zeitliche Zuordnung ist im Diagramm (d) nicht mehr vorhanden. Im tonischen Bereich ist die Punktwolke des Akkordes C-Dur zu sehen. Die Punktwolken der Akkorde F-Dur und G-Dur liegen jeweils im subdominantischen und dominantischen Bereich.

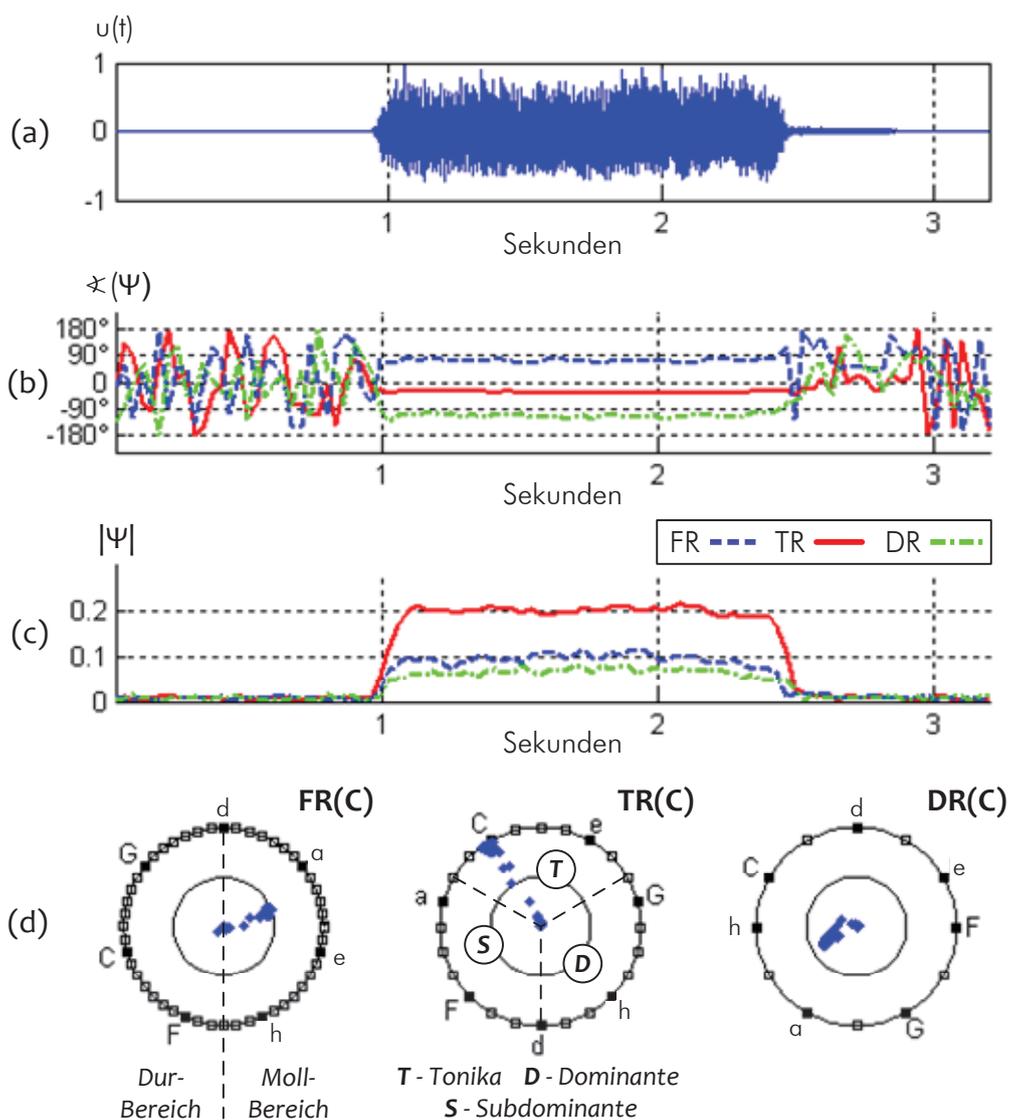


Abbildung 10.6 – Akkord *a*-Moll, Orgel-Klang; (a) Signal im zeitlichen Verlauf; (b) Winkelverlauf in FR, TR und DR; (c) Betragsverlauf in FR, TR und DR; (d) Punktwolken in FR, TR und DR

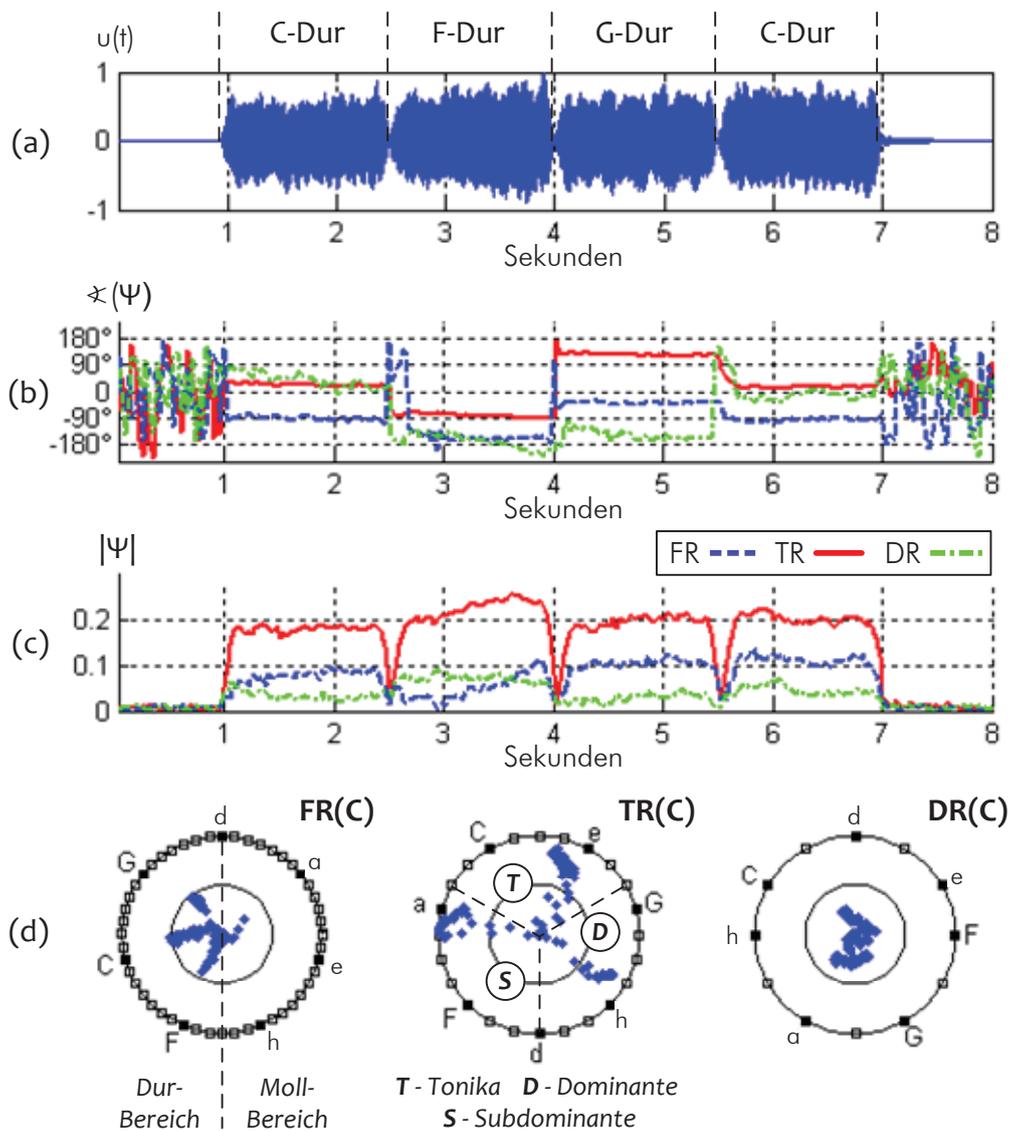


Abbildung 10.7 – Kadenz in C-Dur, Orgel-Klang; (a) Signal im zeitlichen Verlauf; (b) Winkelverlauf in FR, TR und DR; (c) Betragsverlauf in FR, TR und DR; (d) Punktwolken in FR, TR und DR

10.2 Repräsentation von Musikstücken

Dieser Abschnitt zeigt, wie das Symmetriemodell und seine Kreise genutzt werden können, um ein Musikstück zu analysieren. Dazu werden die bekannten Kompositionen „C-Dur Präludium“ von Johann Sebastian Bach, „Für Elise“ von Ludwig van Beethoven und „Eine kleine Nachtmusik“ von Wolfgang Amadeus Mozart näher betrachtet. Die ersten beiden Stücke sind Klavierinterpretationen, während das dritte Stück von einem kleinen Streichorchester gespielt wird. Es werden jeweils die ersten 90 Sekunden der Stücke analysiert und als Punktwolken in den Kreisen T , FR , TR und DR dargestellt. Zudem wird zu den tonartbezogenen Kreisen FR , TR und DR ein Histogramm der Tonigkeitsverteilungen gezeigt, das angibt, wie oft welche Tonigkeit (als Summenvektor) gespielt worden ist. Sowohl die Punktwolken als auch die Histogramme sind auf Grund ihrer Beschaffenheit ohne zeitlichen Bezug.

10.2.1 J. S. Bach „C-Dur Präludium“, BWV 846

Das „C-Dur Präludium“ ist neben der „d-Moll Toccata und Fuge“ (BWV 565) eines der bekanntesten Stücke von Johann Sebastian Bach. Es ist im „Wohltemperierten Klavier Band I“ zu finden. Abbildung 10.8 zeigt die Ausgabe des Symmetriemodells für die ersten 90 Sekunden. Im $TR(C)$ ist zu sehen, dass die drei funktionalen Bereiche Tonika, Dominante und Subdominante in etwa gleichhäufig verwendet werden, wobei der tonische Bereich am meisten gespielt wird. Dies spricht für den ausgewogenen harmonischen Gesamteindruck, welcher vom Stück vermittelt wird. Der $FR(C)$ gibt an, dass hauptsächlich Dur-Akkorde erklingen. Die Histogramm-Verteilung des $DR(C)$ ist nahezu symmetrisch um den Symmetrieton angeordnet.

10.2.2 L. van Beethoven „Für Elise“

Im Gegensatz zum „C-Dur Präludium“ fällt in Abbildung 10.9 auf, dass bei „Für Elise“ die Tonigkeiten im $DR(C)$ nicht symmetrisch um den Symmetrieton verteilt sind. Die Tonigkeiten e und dis kommen weitaus häufiger vor als alle anderen Tonigkeiten. Dies liegt am markanten Beginn des Hauptthemas, das mehrmals erklingt und die Tonigkeiten e und dis abwechselnd verwendet. Am Histogramm des $TR(C)$ ist zu sehen, dass fast ausschließlich sowohl der tonische als auch der dominantische Bereich gespielt werden, während der subdominantische Bereich so gut wie gar nicht erklingt. Dies ist ebenso durch das Hauptthema begründet, da hier nur die Akkorde a -Moll (Tonika) und E -Dur (Dominante) verwendet werden. Das Stück hat einen klaren Moll-Charakter, was im $FR(C)$ und dessen Histogramm-Verteilung zu erkennen ist.

10.2.3 W. A. Mozart „Eine kleine Nachtmusik“, KV 525

Die ebenfalls berühmte Komposition „Eine kleine Nachtmusik“ von W. A. Mozart, die eigentlich den Titel „Serenade in G“ trägt, ist ähnlich strukturiert wie das „C-Dur Präludium“ von J. S. Bach. Abbildung 10.10 zeigt die Ausgabeparameter des Symmetriemodells für das Stück von W. A. Mozart. Es verteilt sich nicht ganz gleichmäßig in den Kreisen des Symmetriemodells, weist aber die gleichen Merkmale auf. Die funktionalen Bereiche Tonika, Dominante und Subdominante werden in etwa gleichhäufig verwendet, wobei Tonika und Dominante leicht überwiegen. Dies zeigt der $TR(G)$ und das spiegelt sich auch in den ersten Takten des Hauptthemas wieder, das durch die Akkorde G-Dur (Tonika) und D-Dur (Dominante) harmonisch begleitet wird. Das Stück hat dadurch einen klaren Dur-Charakter, was sich im $FR(G)$ zeigt. Die Tonigkeitsverteilung im $DR(G)$ ist wesentlich gleichmäßiger als bei „Für Elise“ von L. van Beethoven, jedoch nicht ganz so symmetrisch wie beim „C-Dur Präludium“ von J. S. Bach.

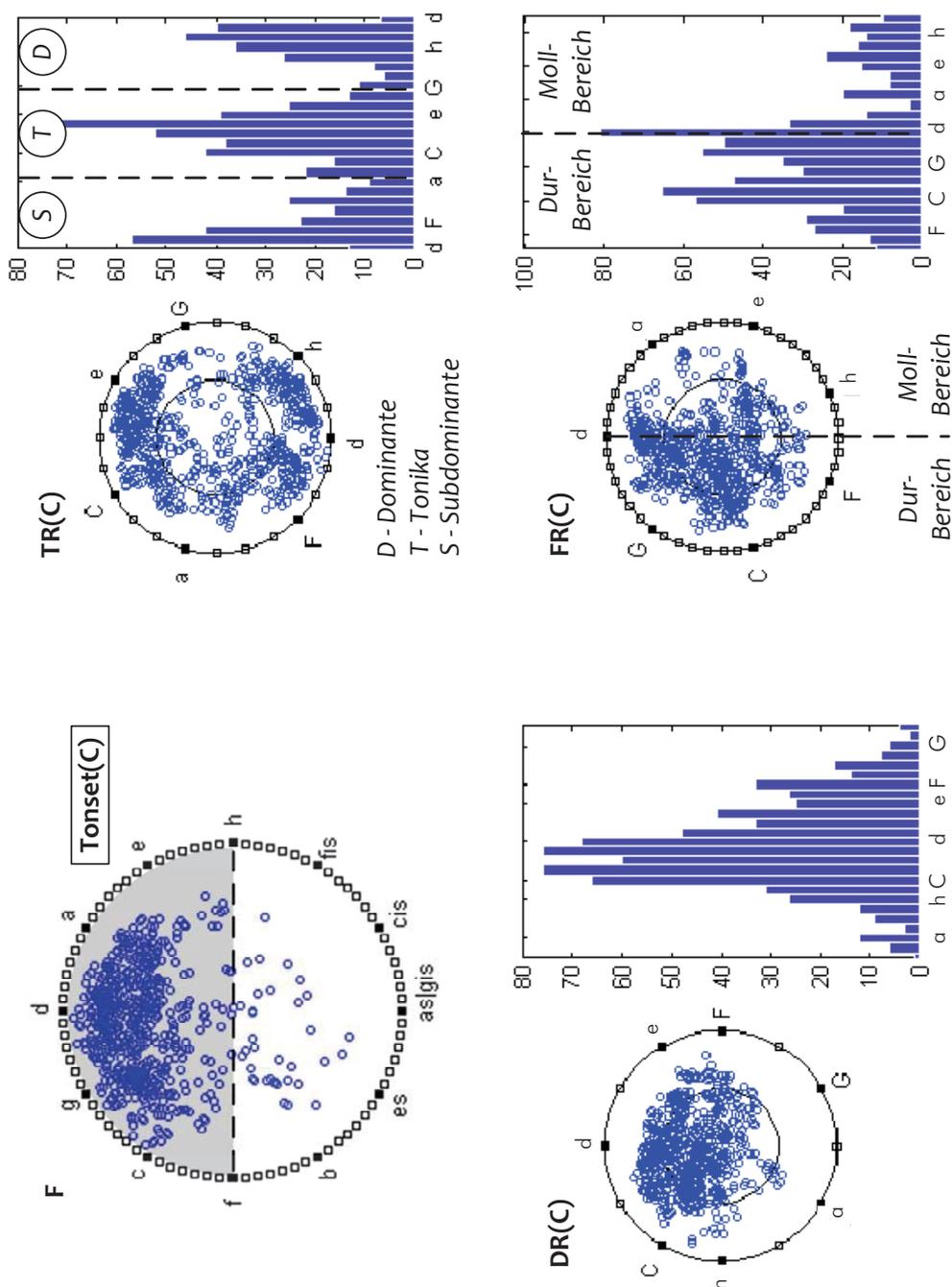


Abbildung 10.8 – J. S. Bach „C-Dur Präludium“, 90 sek., Histogramme und Punktwellen

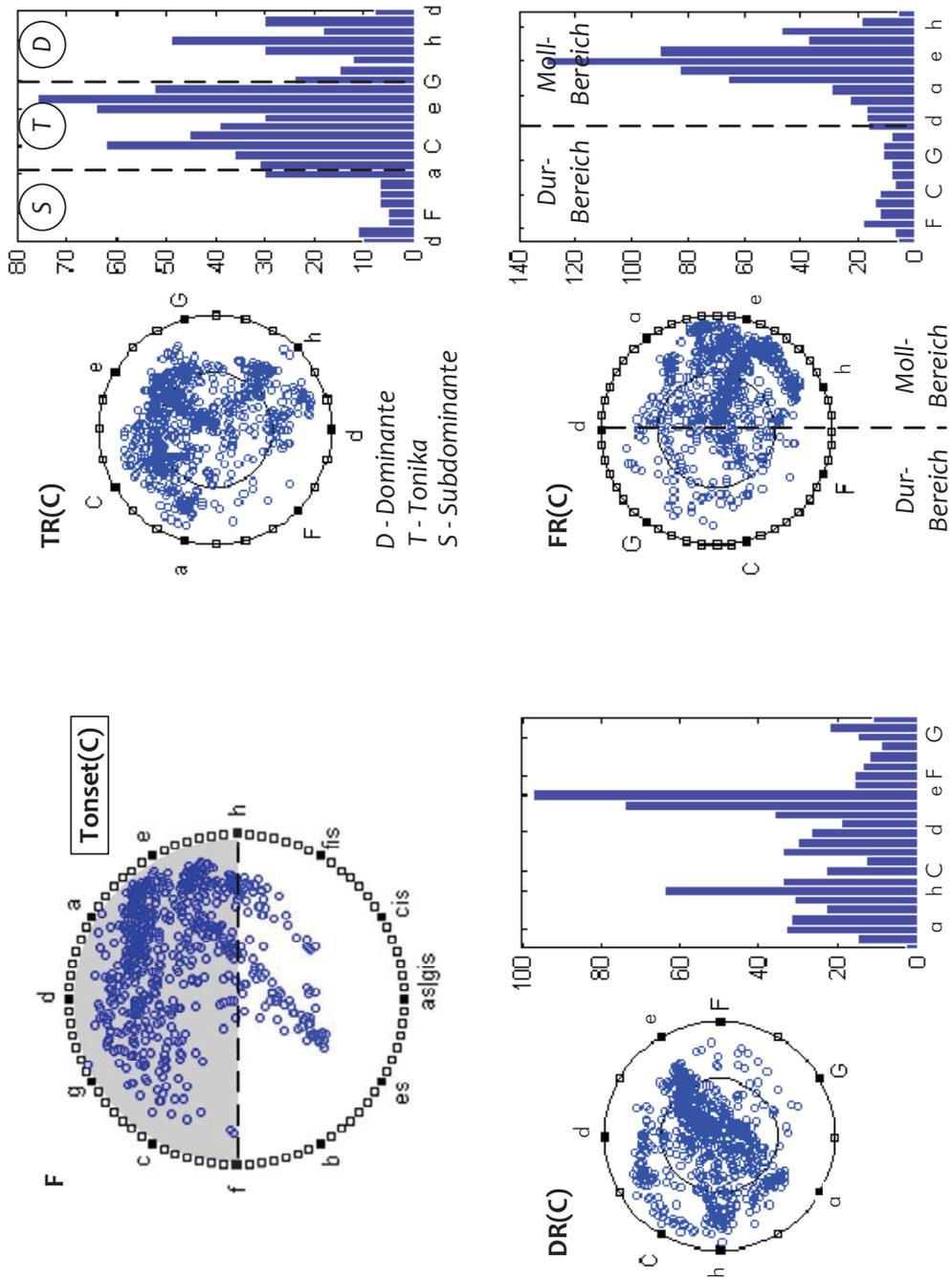


Abbildung 10.9 – L. van Beethoven „Für Elise“, 90 sek., Histogramme und Punktwolken

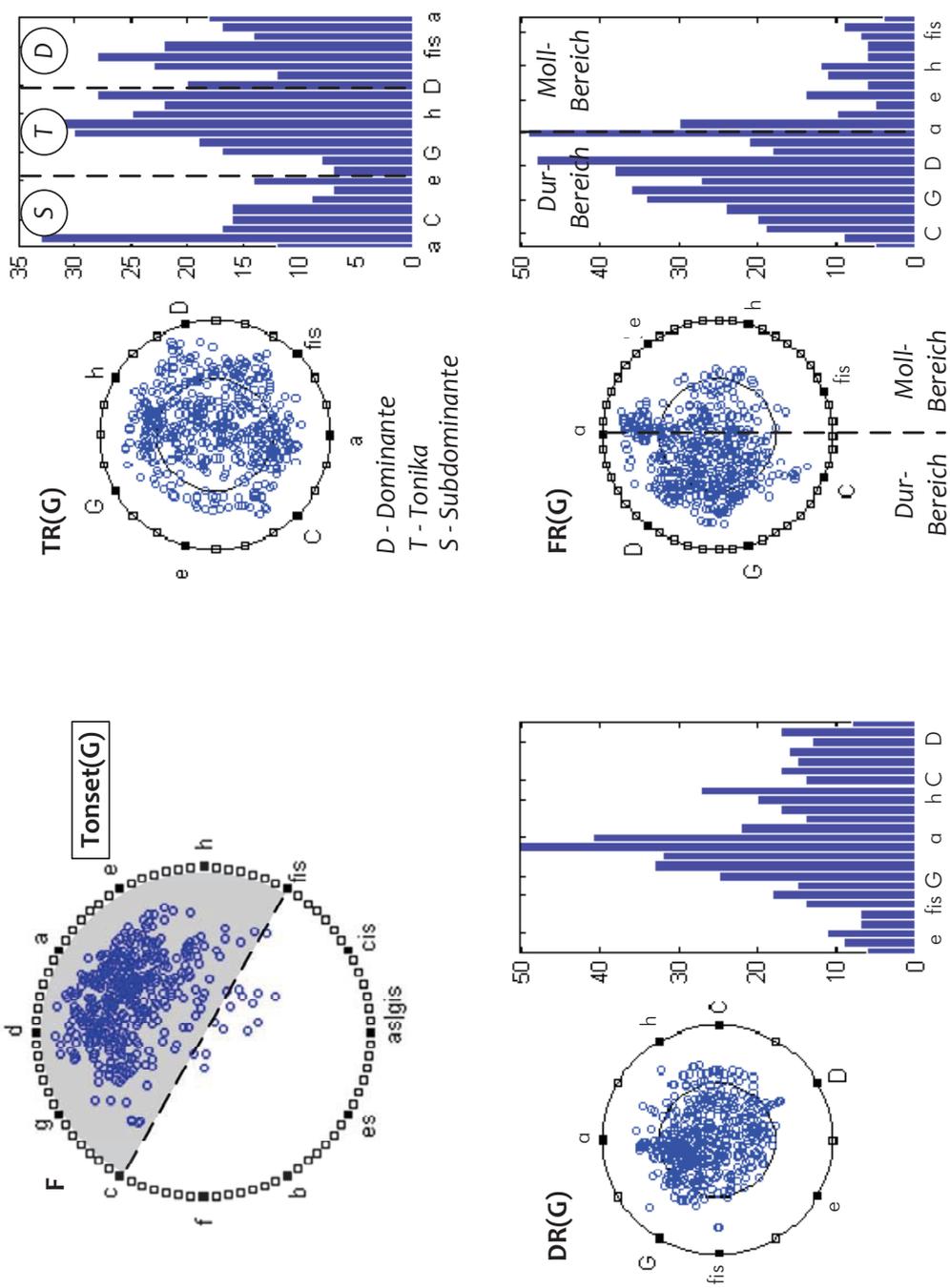


Abbildung 10.10 – W. A. Mozart „Eine kleine Nachtmusik“, 90 sek., Histogramme und Punktwolken

10.3 Zusammenfassung

Die verschiedenen Kreise des Symmetriemodells sind geeignet, musikalische Ereignisse zu analysieren und darzustellen. Dabei kommt den tonartbezogenen Kreisen *FR* und *TR* bei der harmonischen Analyse eine besondere Bedeutung zu: Der *TR* bildet die harmonischen Ereignisse auf die musikalischen Funktionen ab und gruppiert ähnliche harmonische Ereignisse. Der *FR* trennt Dur- und Mollakkorde voneinander und eignet sich so für die Bestimmung des Akkordgeschlechts.

Für die Zuordnung in den Kreisen ist der **Winkel** des Summenvektors ausschlaggebend. Die Länge des Summenvektors gibt zwar die Deutlichkeit des Ereignisses an, jedoch spielt die Länge in einem gewissen Rahmen für die harmonische Analyse keine Rolle. Als Beispiel werden hierfür ein Orgelklang und ein verzerrter E-Gitarren-Klang genannt, die beide den gleichen Akkord spielen. Beide Summenvektoren haben den gleichen Winkel, jedoch stark unterschiedliche Beträge, da das verzerrte E-Gitarren-Signal harmonisches Rauschen beinhaltet, welches die Summenvektorenlänge verkürzt. Diese Erkenntnis zeigt sich auch bei der Akkorderkennung in Kapitel 12. Hier werden die besten Ergebnisse erzielt, wenn man die Winkel der Symmetriemodell-Kreise der *TR* und *FR* als Merkmalsvektoren verwendet. Da bei der Tonarterkennung ebenso die harmonische Struktur des Musikstückes analysiert wird, werden auch hierbei *TR* und *FR* verwendet.

Der tonartübergreifende Quintkreis *F* ist geeignet, den verwendeten Tonigkeitsbereich auszuwählen. So können auf Grund der Lage der Summenvektoren im *F* bestimmte Tonarten und damit auch Tonigkeiten für die weitere Analyse ausgeschlossen werden (vgl. Abbildungen 10.8-10.10). Dadurch wird die Analyse vereinfacht.

Die Eigenschaften des tonartbezogenen diatonischen Kreises *DR* sind für die harmonische Analyse nur bedingt verwertbar, da im *DR* keine harmonischen Ereignisse angezeigt und interpretiert werden können. Da der *DR* aber auf Tonleiterenebene musikalische Ereignisse gruppiert und anordnet, ist er besonders für die Melodieanalyse und die damit verbundenen Algorithmen geeignet.

Kapitel 11

Tonarterkennung

In diesem Kapitel werden zwei neue Algorithmen zur Tonarterkennung vorgestellt. Im ersten Abschnitt wird der Algorithmus zur Erkennung von Dur- und Molltonarten und im zweiten Abschnitt der Algorithmus zur Erkennung von Kirchentonarten eingeführt. Die Evaluierung wird anhand von 100 „Beatles“-Liedern (*Beatles*), 90 Rock/Pop-Stücken (*RockPop*) und 252 klassischen Stücke (*Klassik*) durchgeführt.¹ Zum Training und zum Testen werden jeweils alle 442 Musikstücke verwendet. Auf Grund der großen Anzahl an Trainingsstücken mit verschiedenen Musikstilen führt das Training zu einer Generalisierung.² Der Algorithmus zur Erkennung von Kirchentonarten geht über den Stand der Technik hinaus, da bisher nur zwischen Dur und Moll unterschieden wird.

11.1 Erkennung von Dur- und Molltonarten

In diesem Abschnitt wird ein neuer Algorithmus zur Erkennung von Dur- und Molltonarten vorgestellt. Zunächst wird der Algorithmus selbst erklärt. Anschließend folgt eine Evaluierung mit Musikstücken aus den oben genannten Musikrichtungen *Beatles*, *RockPop* und *Klassik*. Es werden jeweils die ersten 30 Sekunden jedes Stückes analysiert. Die Ergebnisse werden mit einigen Algorithmen aus dem Stand der Technik, u.a von Özgür İzmirli [72] und Alexander Lerch [100] verglichen. Zur Berechnung der Ergebnisse wird der sogenannte MIREX-Score verwendet, der auch Teiltreffer – wie z.B. eine erkannte Nachbarart – in die Ergebnisberechnung mit einfließen lässt. Der MIREX-Score wird in Abschnitt 11.1.2.1 eingeführt.

¹ Die einzelnen Lieder und deren Tonarten aus den Kategorien *Beatles* und *RockPop* sind in Anhang E zu finden.

² Da die Algorithmen sehr wenige Parameter besitzen, kann eine Überanpassung auf die Musikstücke ausgeschlossen werden.

11.1.1 Algorithmus

Abbildung 11.1 zeigt das Blockschaltbild der Dur/Moll-Tonarterkennung. Der Algorithmus ist in zwei Bereiche/Schritte unterteilt:

1. Tonset-Bestimmung,
2. Tongeschlecht-Bestimmung.

Das Tonset (diatonisches Set) ergibt zusammen mit dem Tongeschlecht die zu erkennende Tonart. Das sogenannte *Tonsetprofil* (Bezeichner D' in Abbildung 11.1) ist ein 12-dimensionaler Vektor, der angibt, wie wahrscheinlich zu einem betrachteten Zeitpunkt die zwölf verschiedenen Tonsets sind. Er stellt damit auch dar, welches Tonset nach 30 Sekunden das wahrscheinlichste ist. Eine 30-sekündige Zeitspanne ist in zweifacher Hinsicht vorteilhaft und wird deshalb auch in der Literatur häufig verwendet: 30 Sekunden sind lang genug, um eine Tonart zu etablieren, gleichzeitig sind sie aber so kurz, dass die Wahrscheinlichkeit für einen Tonartwechsel innerhalb dieses Intervalls gering ist.

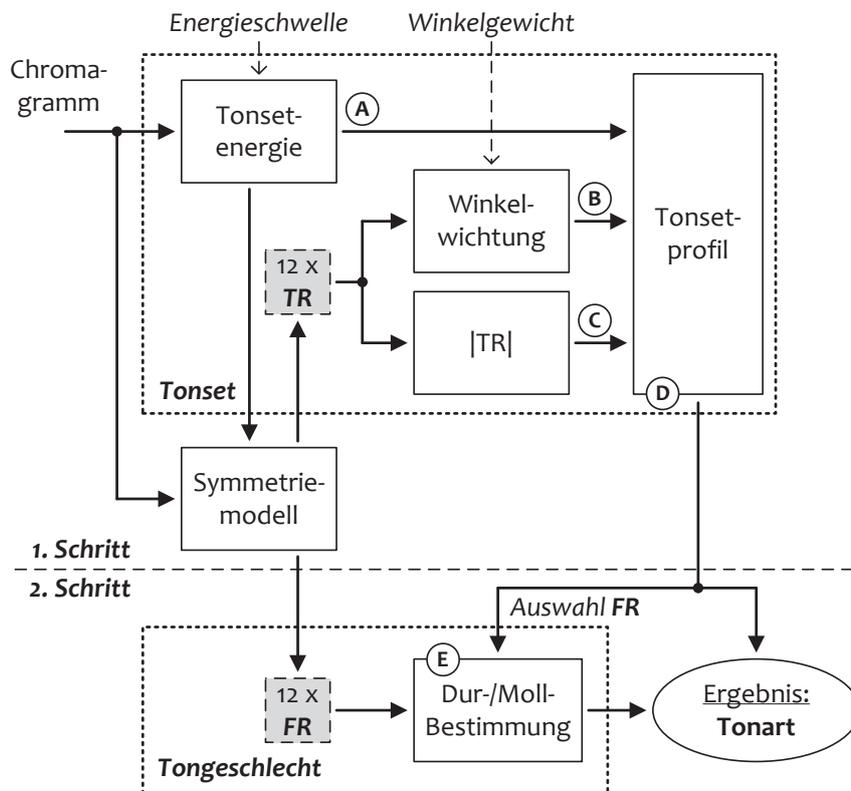


Abbildung 11.1 – Blockschaltbild Tonarterkennung Dur/Moll; (A): Tonsetenergie, (B): Winkel-wichtung, (C): Summenvektorbetrag im TR , (D): Tonsetprofil, (E): Dur-/Moll-Bestimmung im FR

Abbildung 11.2 zeigt beispielhaft das Tonsetprofil für das Stück „Any time at all“ von den Beatles, dessen Tonart *D-Dur* ist, nach jeweils 10 s, 20 s und 30 s. Das Tonset mit der größten Wahrscheinlichkeit wird nach 30 s als Tonset ausgewählt und gibt vor, in welchem *FR* (vgl. Abbildung 11.1) die Ermittlung des Tongeschlechts durchgeführt wird. Für das gezeigte Beispiel ist dies das Tonset mit den Tonigkeiten von *D-Dur*.

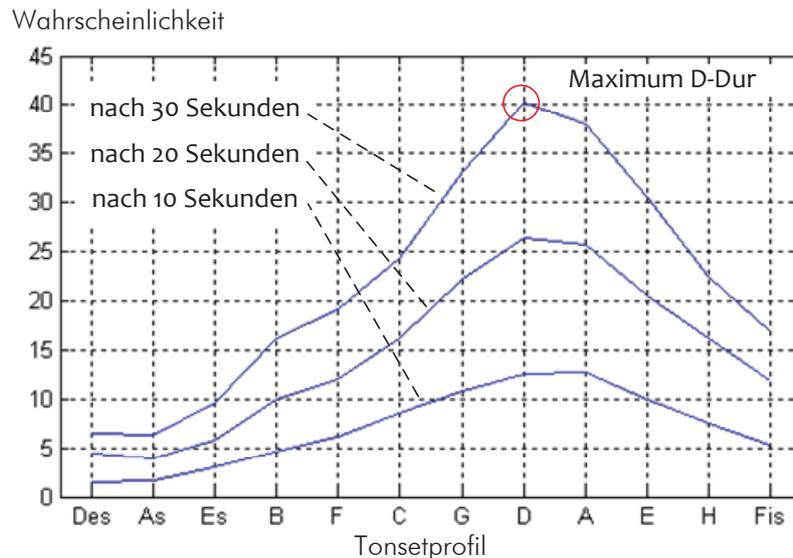


Abbildung 11.2 – Tonsetprofil; „Any time at all“, Beatles

Die Wahrscheinlichkeiten für das Tonsetprofil werden aus dem Produkt von bis zu drei verschiedenen Einzelwahrscheinlichkeiten berechnet, der *Tonsetenergie* (*A*), der *Winkelwichtung* (*B*) und den *Beträgen in TR* (*C*) (vgl. Abbildung 11.1). Diese drei Wahrscheinlichkeitsmaße werden im Folgenden näher betrachtet.

11.1.1.1 Tonsetenergie

Ein Chromavektor beinhaltet zwölf Tonigkeiten. Jedes diatonische Set besteht aus sieben dieser zwölf Tonigkeiten. Je häufiger eine Tonigkeit eines Tonsets im Musikstück gespielt wird, desto wahrscheinlicher ist die Etablierung dieses Tonsets. Die relative Tonsetenergie $E(k)$ bezogen auf die Gesamtenergie eines Chromavektors gibt an, wie wahrscheinlich ein Tonset k für das aktuelle Analysefenster ist:

$$E(k) = \frac{\sum_{n_c=-6}^5 (\varepsilon_{n_c} * z)^2}{\sum_{n_c=-6}^5 (\varepsilon_{n_c})^2} \quad (11.1)$$

$z = 1$ | n_c ist die Tonigkeit des Tonsets k

$z = 0$ | n_c ist nicht Tonigkeit des Tonsets k

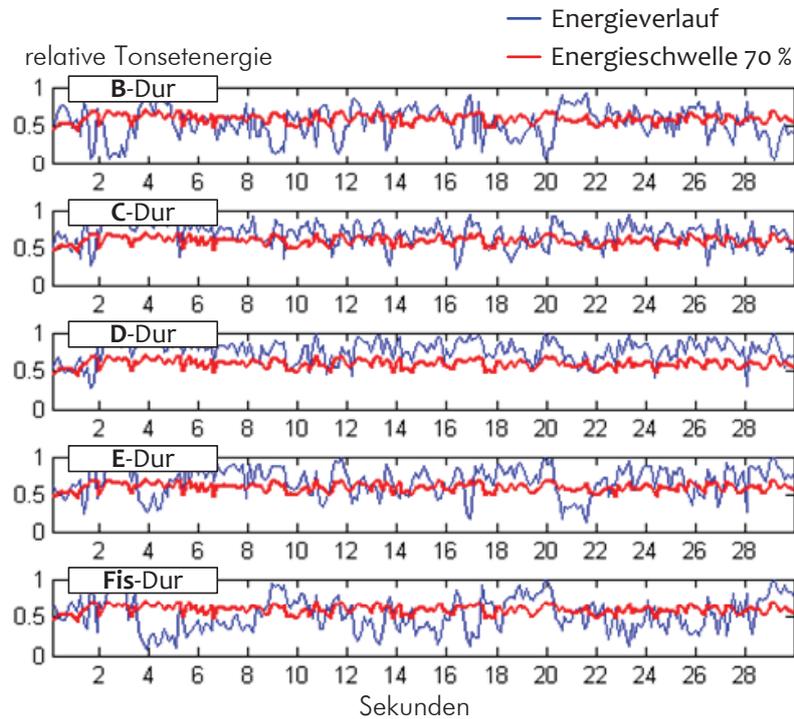


Abbildung 11.3 – Tonsetenergie und Energieschwelle; „Any time at all“, Beatles

In Gleichung 11.1 steht ε_{n_c} für die Amplitude der entsprechenden Tonigkeit. Abbildung 11.3 zeigt den Verlauf der Tonsetenergie für fünf ausgewählte Tonsets des betrachteten Musikstücks „Any time at all“. Die Energieschwelle ist eine Kurve, die einen zu evaluierenden Prozentsatz unterhalb der relativen Maximalenergie liegt. Oberhalb der Kurve werden die Wahrscheinlichkeiten der einzelnen Tonsets für die Ermittlung des Tonsetprofils verwendet. Unterhalb der Kurve werden die Wahrscheinlichkeiten zu Null gesetzt. Die Summenvektoren in den dazugehörigen *FR* werden für diese Zeitfenster verworfen und für die Tongeschlechtsbestimmung nicht verwendet, da keine harmonisch sinnvollen Ereignisse in diesen *FR* enthalten sind. Der optimale Wert der Energieschwelle wird in Abschnitt 11.1.2 evaluiert. Im Beispiel wird zur Veranschaulichung eine Energieschwelle von 70 % verwendet. In Abbildung 11.3 ist zu sehen, dass für das Tonset von *D-Dur* die relative Tonsetenergie fast immer über der Schwelle liegt. Damit ist das Tonset von *D-Dur* sehr wahrscheinlich das richtige Tonset. Für die Tonsets von *B-Dur* und *Fis-Dur* befinden sich die relativen Tonsetenergien meist unterhalb der Energieschwelle.

11.1.1.2 Winkelwichtung

Jeder Dreiklang ist in drei verschiedenen diatonischen Tonräumen vertreten, einmal als Tonika, einmal als Dominante und einmal als Subdominante: z.B. ist der Akkord *C-Dur*

in der Tonart C-Dur die Tonika, in F-Dur die Dominante und in G-Dur die Subdominante. Dies ist in Abbildung 11.4 zu sehen. Die Punktwolken entsprechen einem gespielten C-Dur-Akkord.

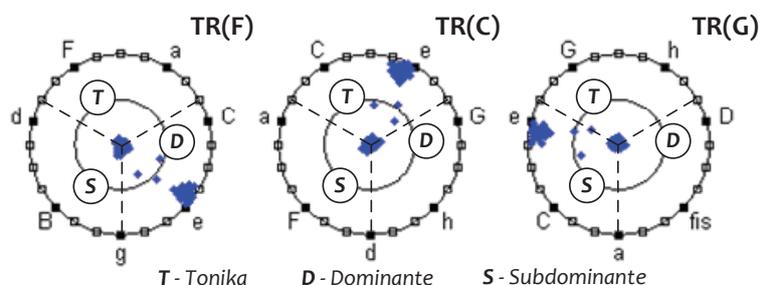


Abbildung 11.4 – C-Dur als Dominante, Tonika und Subdominante

Der Summenvektorwinkel der **Tonika** ist im *TR* am kleinsten, weil hier der tonische Bereich oben im Kreis liegt. Es wird nach demjenigen diatonischen Set gesucht, in dem der am häufigsten gespielte Akkord der Tonika zugeordnet wird. Daher kann der Winkel als Bewertungsmaß für die Ermittlung des richtigen Tonsets miteinbezogen werden. Dazu wird in Abhängigkeit des Summenvektorwinkels im *TR* eine Wahrscheinlichkeitsfunktion, die sogenannte *Winkelwichtung*, definiert, die mit größer werdendem Betrag des Summenvektorwinkels eine kleinere Wahrscheinlichkeit für den betrachteten *TR* und damit für das diatonische Set ausgibt. Diese Funktion wird in Abbildung 11.5 beispielhaft mit einem Winkelgewicht von 0,6 dargestellt.

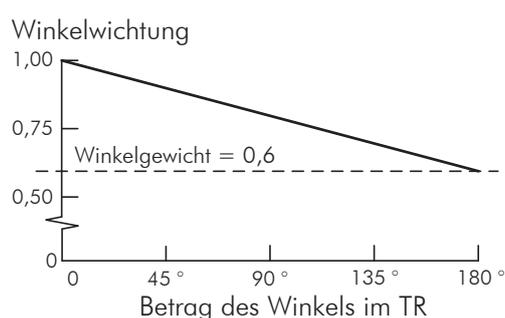


Abbildung 11.5 – Winkelwichtung; Winkelgewicht = 0,6

Das *Winkelgewicht* legt fest, welche Wahrscheinlichkeit für den größten Winkelbetrag 180° und damit für das von der Tonika am weitesten entfernt liegenden Ereignis ausgegeben wird. Der Wert des optimalen Winkelgewichts wird in Abschnitt 11.1.2 evaluiert. Die Winkelwichtung für das betrachtete Musikstück „Any time at all“ wird für fünf ausgewählte diatonische Sets in Abbildung 11.6 gezeigt. Anhand des Verlaufs für das Tonset

von *D-Dur* ist zu sehen, dass hier die Winkelwichtung häufiger als die Winkelwichtung der anderen Tonsets einen Wert nahe 1,0 hat.

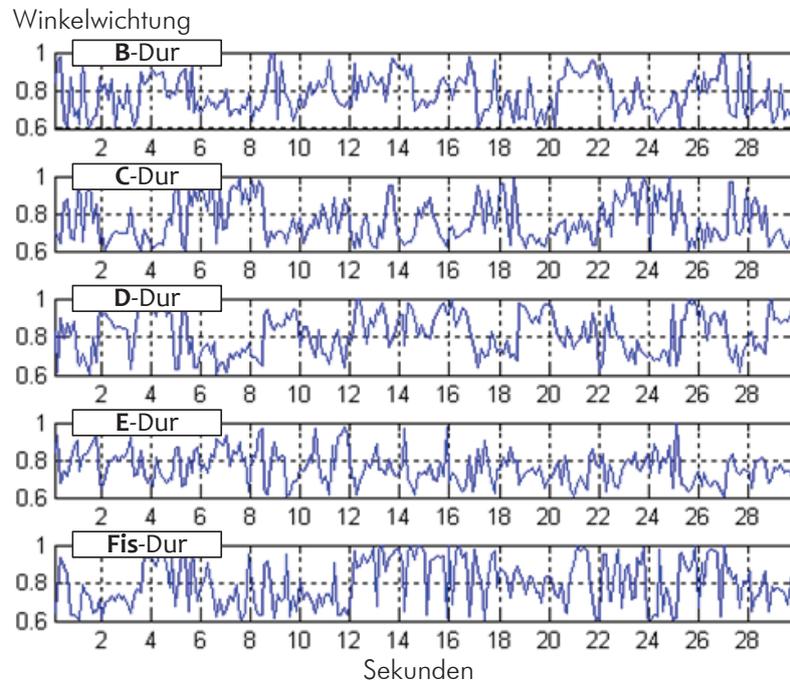


Abbildung 11.6 – Winkelwichtung; Winkelgewicht = 0,6; „Any time at all“, Beatles

11.1.1.3 Summenvektorbetrag in TR

Der Summenvektorbetrag in einem *TR* gibt an, wie stark das harmonische Ereignis in diesem ausgeprägt ist. Da die Dreiklänge Grundlage aller Harmoniebildung sind, liegt es nahe, in den *TR* nach langen Summenvektoren zu suchen. In dieser Tonraumart sind im Vergleich zu den *FR* und *DR* die Summenvektoren für Dreiklänge am längsten (vgl. Kapitel 10). Abbildung 11.7 stellt den auf den Maximalwert normierten Betragsverlauf im *TR* für fünf ausgewählte diatonische Sets und das betrachtete Musikbeispiel dar. In Abschnitt 11.1.2 wird untersucht, ob die Summenvektorbeträge für die Bestimmung des Tonsetprofils geeignet sind.

11.1.1.4 Ermittlung des Tongeschlechts

Nach Ermittlung des richtigen Tonsets wird im dazugehörigen *FR* bestimmt, ob das Stück in Dur oder Moll gespielt wird (Bezeichner ‚E‘ in Abbildung 11.1). Dazu werden die Summenvektorpunkte auf der linken und rechten Seite des *FR* ausgezählt. Dies ist gleichbedeutend mit der Beantwortung der Frage, ob im Stück häufiger Dur- oder Moll-

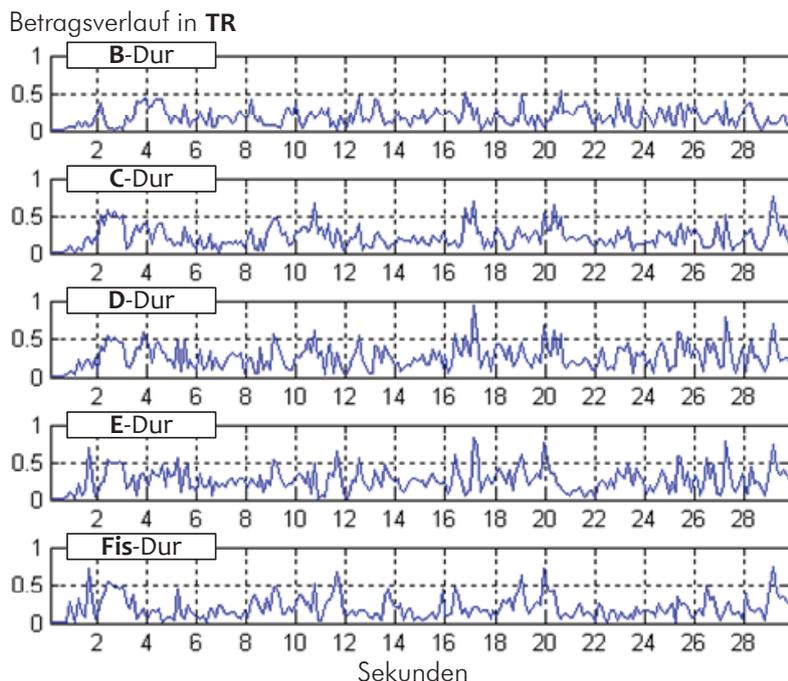


Abbildung 11.7 – Betragsverlauf in TR; „Any time at all“, Beatles

Harmonien erklingen. Für das betrachtete Musikstück „Any time at all“ von den Beatles ergibt sich (vgl. Abbildung 11.8), dass mehr Summenvektoren im Dur- als im Moll-Bereich liegen. Damit ist die Tonart des zu analysierenden Musikstückes *D-Dur*.

11.1.2 Evaluierung

Der vorgestellte Algorithmus wird in diesem Abschnitt evaluiert, d.h. es wird ermittelt, welche Einzelwahrscheinlichkeiten (vgl. Abbildung 11.1) für die Tonsetbestimmung und damit auch für die Dur/Moll-Tonarerkennung geeignet sind. Dazu werden die *Tonsetenergie* (*A*), die *Winkelwichtung* (*B*) und die *Beträge in TR* (*C*) unterschiedlich kombiniert und gleichzeitig diejenigen Werte für die *Energieschwelle* und das *Winkelgewicht* bestimmt, die bei gegebenem Trainingsdatensatz zu einer maximalen Tonarterkennungsrate führen. Die Tonarterkennungsrate wird mit dem sogenannten MIREX-Score ermittelt, der nachfolgend erläutert wird.

11.1.2.1 MIREX-Score

Der MIREX-Score wurde 2005 zur Auswertung und zum Vergleich von Tonarterkennungsalgorithmen eingeführt. Neben der Erkennung der korrekten Tonart werden für „Nachbartonarten“ Teiltreffer vergeben (vgl. [72]). Zu den Nachbartonarten zählen ers-

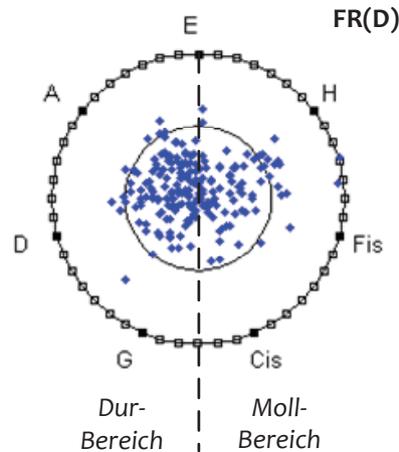


Abbildung 11.8 – Dur/Moll-Bestimmung; „Any time at all“, Beatles

tens die Tonarten, die eine Quinte höher und tiefer liegen, zweitens die parallelen Dur/Moll-Tonarten und drittens die relative Tonart, d.h. wenn z.B. anstelle von C-Dur die Tonart c-Moll erkannt wird. Tabelle 11.1 zeigt den Bewertungsmaßstab für die Dur/Moll-Erkennung des MIREX-Scores. Die Tonarterkennungsrate wird berechnet, indem die Summe der Trefferpunkte durch die Gesamtanzahl aller Teststücke dividiert wird.

Tonart	Trefferpunkte
Korrekte Tonart	1,0
Quintverwandte Tonart	0,5
Parallele Tonart	0,3
Relative Tonart	0,2

Tabelle 11.1 – MIREX-Score

11.1.2.2 Parameter-Evaluierung

Die Wirkung und Optimierung des Parameters Energieschwelle zeigt Abbildung 11.9. Der Parameter Winkelgewicht bleibt dabei konstant. Als Teststücke werden die Stücke der Kategorie *Beatles* verwendet. Es zeigt sich, dass im Bereich 0,7 bis 0,9 die Tonarterkennungsrate am größten ist und ca. 3 % höher liegt als ohne Energieschwelle (Energieschwelle = 0). Abbildung 11.10 zeigt die Optimierung und den Einfluss des Parameters Winkelgewicht. Der Parameter Energieschwelle bleibt dabei konstant. Es werden dieselben Teststücke wie bei der Evaluierung der Energieschwelle verwendet.

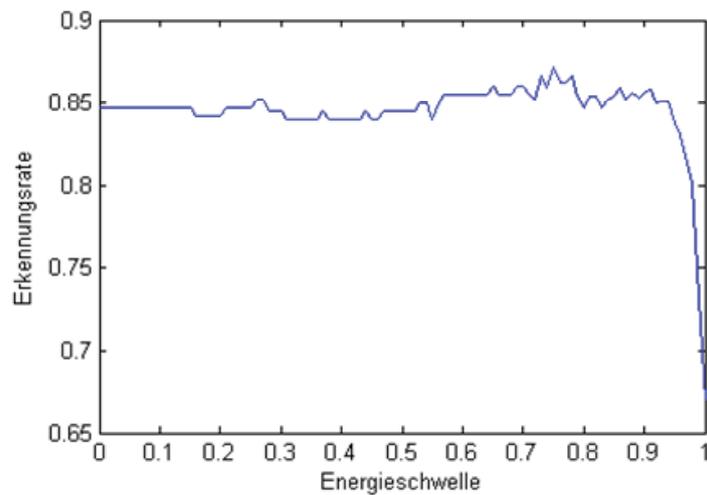


Abbildung 11.9 – Parameteroptimierung Energieschwelle

Das Maximum der Tonarterkennungsrate wird für ein Winkelgewicht im Bereich 0,6 bis 0,8 erzielt und ist knapp 10 % größer als ohne Winkelwichtung (Winkelgewicht = 0).

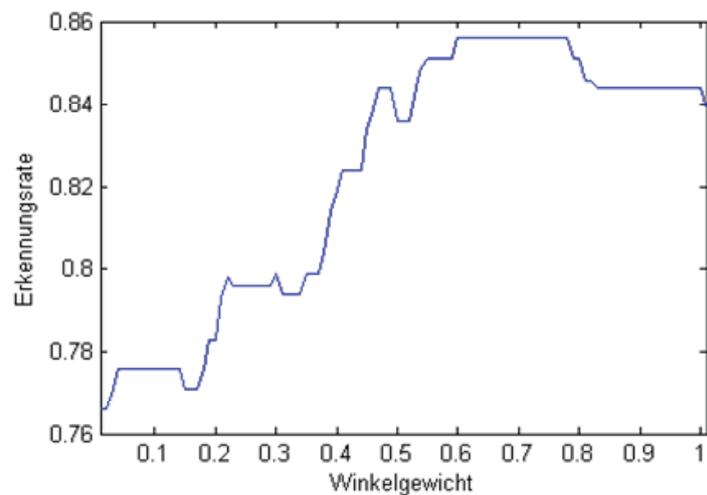


Abbildung 11.10 – Parameteroptimierung Winkelgewicht

11.1.2.3 Ergebnis der Evaluierung

Zur Berechnung der Tonsetwahrscheinlichkeit und Tonarterkennungsrate werden die Kombinationen $A \cdot B \cdot C$, $A \cdot B$ und $A \cdot C$ genutzt und verglichen. Tabelle 11.2 listet die Ergebnisse der Evaluierung auf. Die Kombination aus der *Tonsetenergie* A und *Winkelwichtung* B ergibt die besten Ergebnisse mit einer Erkennungsrate von im Durchschnitt 82,9 %. Das höchste Einzelergebnis mit 87,1 % wird für die Kategorie *Klassik* erzielt.

	$A \cdot B \cdot C$	$A \cdot B$	$A \cdot C$
Energieschwelle	0,92	0,91	0,95
Winkelgewicht	0,62	0,74	nicht verwendet
<i>Beatles</i>	86,1 %	85,9 %	74,1 %
<i>RockPop</i>	77,6 %	75,6 %	65,4 %
<i>Klassik</i>	83,9 %	87,1 %	79,9 %
Durchschnitt	82,5 %	82,9 %	73,1 %

Tabelle 11.2 – Ergebnisse der Dur/Moll-Evaluierung nach den ersten 30 Sekunden; Parameterevaluierung und Tonarterkennungsraten

Die verwendeten Tonarten der Musikstücke in der Kategorie *Klassik* sind ausschließlich die Tonarten Ionisch und Aeolisch. Die Winkelwichtung bevorzugt aufgrund ihres Aufbaus diese beiden Tonarten und benachteiligt die anderen vier Kirchentonarten Dorisch, Phrygisch, Lydisch und Mixolydisch. Bei der Bestimmung des richtigen Tonsets wird dadurch sehr oft dasjenige Tonset ausgewählt, bei dem der Grundakkord des Musikstücks, die Tonika, zur Tonart Ionisch oder Aeolisch zugeordnet wird. Wenn aber z.B. die richtige Tonart eines Musikstückes Dorisch wäre, so würden im ausgewählten Tonset, der den Grundakkord zu Aeolisch zuordnet, einige Tonigkeiten fehlen. Dadurch wird die Dur/Moll-Bestimmung in *FR* verfälscht. Dies erklärt auch, dass die Ergebnisse für *RockPop* am niedrigsten sind, da hier prozentual die meisten Tonarten vorkommen, die nicht Ionisch oder Aeolisch sind. Die Ergebnisse für die Kombination aus *Tonsetenergie A* und *Winkelwichtung B* werden für den Vergleich mit anderen Verfahren im nächsten Abschnitt verwendet.

11.1.3 Vergleich mit anderen Verfahren

Zur Bewertung der Leistungsfähigkeit des neu vorgestellten Algorithmus' zur Tonarterkennung wurden weitere Algorithmen, die dem Stand der Technik entsprechen, implementiert. Folgende Algorithmen werden verglichen:

- **Izmirli:** Algorithmus von Özgür İzmirli [72]. Dieser Algorithmus erzielte bei der MIREX2005 die besten Ergebnisse. Die selbst generierten Tonartprofile hat Özgür İzmirli für die Untersuchung zur Verfügung gestellt.

- **Lerch_Original:** Algorithmus von Alexander Lerch [100], der von den Algorithmen aus dem Stand der Technik die besten Ergebnisse erzielt.
- **Zhu:** Algorithmus von Yongweih Zhu und Kollegen [168]. Der verwendete Ansatz zur Tonarterkennung ist ähnlich wie der hier neu vorgestellte Algorithmus zweistufig.
- **Lerch_neu:** Alexander Lerch verwendet in seinem eigenen Ansatz KK-Tonartprofile. Die Implementierung hat ergeben, dass TY-Tonartprofile für die verwendeten Musikstücke wesentlich bessere Ergebnisse erzielen. Lerch_neu ist damit eine Abwandlung des originalen Algorithmus’.
- **SYM_Modell:** Dieser Algorithmus ist der hier neu vorgestellte Algorithmus zur Tonarterkennung, der auf dem Symmetriemodell basiert.
- **DurMoll:** Bei diesem Algorithmus wird der Mittelwert aus allen Chromavektoren gebildet, der dann mit den 24 möglichen Dur/Moll-Dreiklängen korreliert wird. Der Akkord, für den die höchste Korrelation erzielt wird, gibt die erkannte Tonart vor. Emilia Gomez [51] hat bereits darauf hingewiesen, dass einfache Dur/Moll-Dreiklänge als Tonartprofile für bestimmte Musikrichtungen besser geeignet sein können als KK- und TY-Tonartprofile.

Die Einzelergebnisse für die Kategorien *Beatles*, *RockPop* und *Klassik* sowie das Gesamtergebnis sind in Tabelle 11.3 aufgelistet mit dem MIREX-Score als Bewertungsmaßstab. Der neue Algorithmus SYM_Modell erzielt für alle Kategorien das beste Ergebnis. Für *Beatles* beträgt die Erkennungsrate 85,9 %, für *RockPop* 75,6 % und für *Klassik* 87,1 %; somit ist der Durchschnitt 82,9 % an richtig erkannten Tonarten.

Der Algorithmus von İzmirlı schneidet hier mit im Durchschnitt 53,8 % am schlechtesten ab. Eine Begründung hierfür liegt darin, dass die von İzmirlı selbst generierten Tonartprofile aus Klaviertönen erstellt sind und Klaviertöne, z.B. bei den Beatles, nur in wenigen Stücken explizit zu hören sind. Zu der Qualität der Profile merkt Özgür İzmirlı an, dass für jeden Musikstil und jede Instrumentierung neue Tonartprofile zu erstellen seien. Dies ist jedoch nicht praktikabel, da bei der Erkennung von Musik davon ausgegangen werden muss, dass keinerlei Vorwissen über das Stück besteht. Aber genau dies wird beim Ansatz von İzmirlı benötigt. Der Algorithmus von Lerch (Lerch_Original) hat mit 64,2 % die höchste Erkennungsrate des derzeitigen Standes der Technik. Der Algorithmus von Zhu und Kollegen (Zhu) erreicht im Durchschnitt 60,7 %.

Der hier neu vorgestellte Algorithmus (SYM_Modell) ist damit im Mittel knapp 30 % besser als der beste Algorithmus (İzmirlı) der MIREX2005. Bei der MIREX2005 wurden nur

	<i>Beatles</i>	<i>RockPop</i>	<i>Klassik</i>	Durchschnitt	
Izmirli	42,3 %	49,4 %	69,5 %	53,8 %	Stand der Technik
Lerch_Original	60,7 %	68,9 %	63,1 %	64,2 %	
Zhu	60,7 %	53,6 %	67,8 %	60,7 %	
Lerch_neu	82,5 %	73,3 %	77,7 %	77,8 %	Neu
SYM_Modell	85,9 %	75,6 %	87,1 %	82,9 %	
DurMoll	81,8 %	71,6 %	79,4 %	77,6 %	

Tabelle 11.3 – Ergebnisse Tonarterkennung nach MIREX-Score; Mittelwerte

resynthetisierte Musikstücke verwendet. Dies kann erklären, dass die angegebene Erkennungsrate von *İzmirli* bei der MIREX2005 bei 89,5 % lag, da bei resynthetisierter Musik die Tonarterkennung einer symbolischen Methode (auf reinen Noten basierend) näher kommt als die Erkennung von echtem Audio, welches in dieser Arbeit verwendet wird. Beim Algorithmus *Lerch_neu* verbessert sich die Erkennungsrate um 13,6 % auf 77,8 %. Das war nicht zu erwarten, da die KK- und TY-Tonartprofile sich zwar unterscheiden, in der Grundstruktur aber dennoch sehr ähnlich sind. Dies zeigt, dass die Verwendung dieser Profile zu nicht vorhersagbaren Ergebnissen führen kann. Die Untersuchungen ergaben, dass selbst bei ähnlichen Musikstücken und Verwendung der gleichen Methode, die auf diesen Profilen aufbaut, die Ergebnisse starken Schwankungen unterworfen sind. Es ist anzumerken, dass sowohl die KK-Tonartprofile als auch die TY-Tonartprofile Musikstücke voraussetzen, deren Tonart entweder Ionisch-Dur oder Aeolisch-Moll ist, was bisher so noch nicht diskutiert worden ist. Dieser Ansatz ist für klassische Stücke durchaus verwendbar, da hier fast nur Ionisch-Dur und Aeolisch-Moll vertreten sind. Dies fällt vor allem beim Algorithmus von *İzmirli* auf. Hier ist die Erkennungsrate für *Klassik* mit Abstand am höchsten. Für Musikstücke anderer Musikepochen, auch der heutigen, sind diese Ansätze eher ungeeignet, da viel häufiger von den Kirchentonarten Gebrauch gemacht wird: So werden Lydisch-Dur und Mixolydisch-Dur sowie Dorisch-Moll und Phrygisch-Moll in der aktuellen Musik aber auch schon von den Beatles verwendet. Hierbei müssen die KK- und TY-Tonartprofile scheitern, da das darin verwendete Tonset nicht zu dem des Stückes passt. Diesen Überlegungen folgend wären sechs Tonartprofile zu erstellen. Allerdings wären sich diese Profile so ähnlich, dass die Erkennung der korrekten Tonart eher zufällig geschehe. Die schlichte Schlussfolgerung ist, dass einfache

Dur/Moll-Tonartprofile, die nur die drei Tonigkeiten der Tonika beinhalten, für moderne Musik besser geeignet sein können. Damit ist Emilia Gomez Annahme bestätigt. Mit diesen einfachen Profilen wird im Prinzip nur nach dem häufigsten Akkord gesucht, was wesentlich sinnvoller ist, als nach dem häufigsten Akkord zu suchen und vorauszusetzen, dass dieser entweder nur die Tonika von Ionisch-Dur oder von Aeolisch-Moll sein kann. Genau dies machen aber die KK- und TY-Tonartprofile. Die Ergebnisse des Algorithmus' DurMoll, der nur Dur/Moll-Tonartprofile verwendet, belegt dies mit einer durchschnittlichen Tonarterkennungsrate von 77,6 %. Jedoch ist dieses Ergebnis immer noch mehr als 5 % unter dem Ergebnis des hier neu vorgestellten Algorithmus'.

Zur statistischen Auswertung der Ergebnisse wird der Friedman-Rangsummen-Test verwendet. Dieser bestimmt, ob sich die berechneten Mittelwerte signifikant unterscheiden. Es wird die übliche Fehlerschranke von 5 % verwendet. Abbildungen 11.11-11.14 zeigen die mittleren Rangsummen (Kreise) und das dazugehörige 95 %-Konfidenzintervall (Linien) gemeinsam für alle Teststücke und jeweils einzeln für die drei Kategorien *Beatles*, *RockPop* und *Klassik*. Der hier neu vorgestellte Algorithmus ist mit einer Sicherheit von über 95 % für alle Musikstücke (vgl. Abbildung 11.11) besser als der Stand der Technik (Izmirli, Lerch_Original, Zhu). Dies gilt auch für den Algorithmus Lerch_neu. Die Konfidenzintervalle des neuen Algorithmus' SYM_Modell und des Algorithmus' DurMoll überlappen sich leicht. Dies bedeutet, dass die Sicherheit hierbei etwas geringer als 95 % ist.

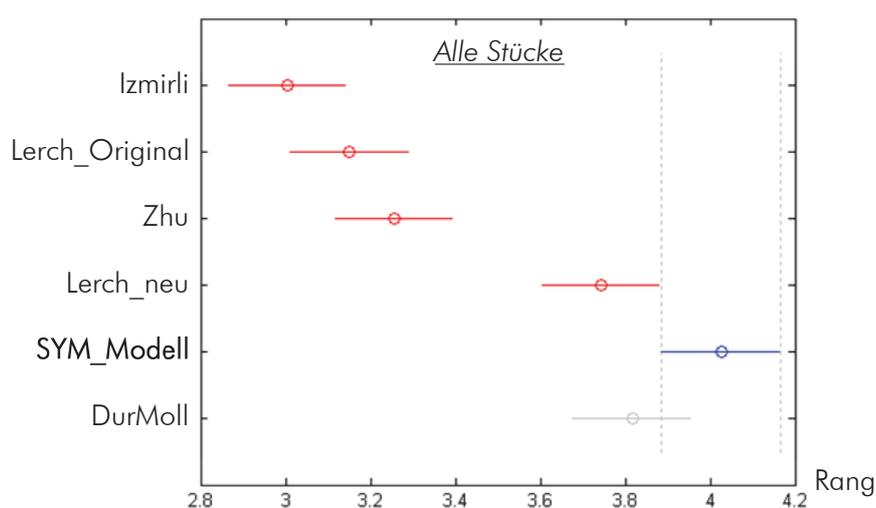


Abbildung 11.11 – Ergebnisse Tonarterkennung; Alle Stücke; Friedman-Test; $\alpha = 5\%$

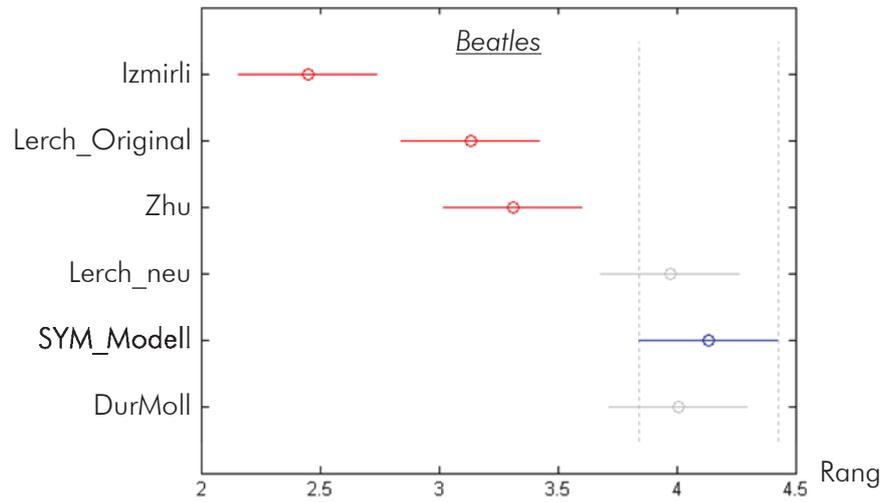


Abbildung 11.12 – Ergebnisse Tonarterkennung; Beatles; Friedman-Test; $\alpha = 5\%$

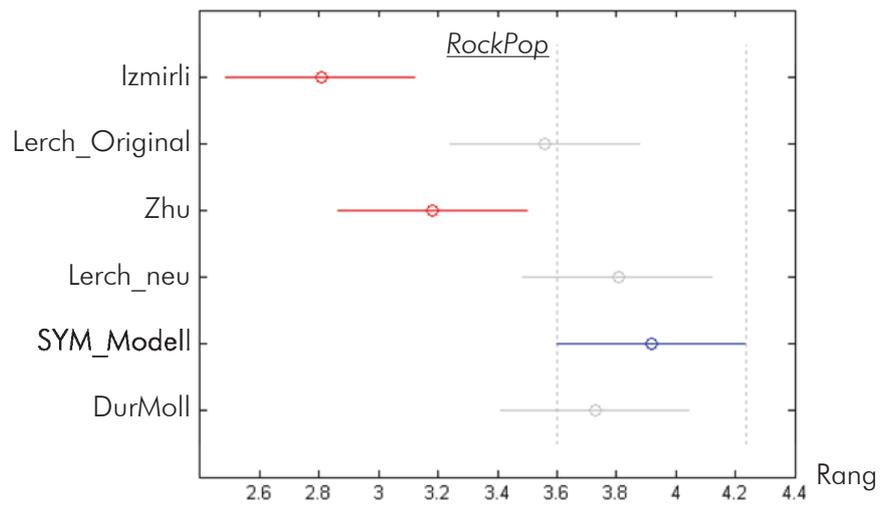


Abbildung 11.13 – Ergebnisse Tonarterkennung; RockPop; Friedman-Test; $\alpha = 5\%$

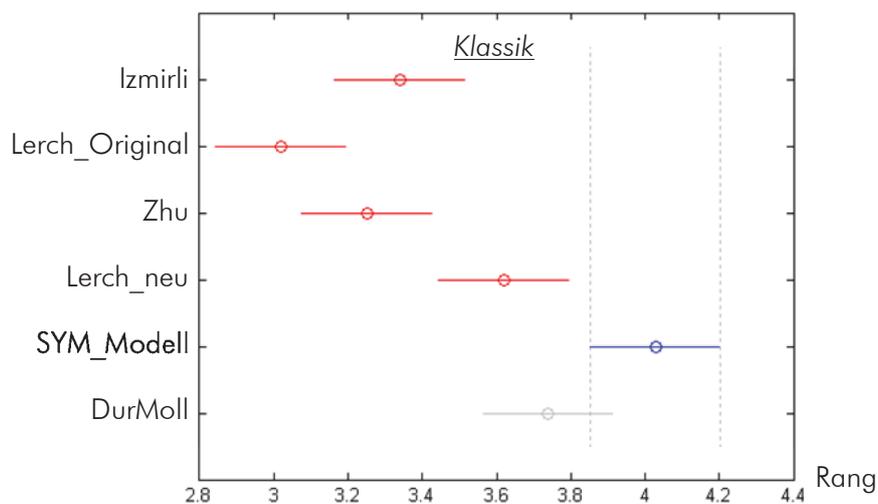


Abbildung 11.14 – Ergebnisse Tonarterkennung; Klassik; Friedman-Test; $\alpha = 5\%$

11.2 Erkennung von Kirchentonarten

Dieser Abschnitt stellt einen neuen Algorithmus zur Tonarterkennung vor. Das Ziel dieses Algorithmus' ist es, die Tonart eines Musikstückes zu analysieren und einer der sechs Kirchentonarten Ionisch, Dorisch, Phrygisch, Lydisch, Mixolydisch oder Aeolisch zuzuordnen. Dieser Ansatz geht über den Stand der Technik hinaus, da in der Musikanalyse bisher nur zwischen Dur und Moll unterschieden wird. Da die Tonart den grundlegenden Charakter eines Stückes wiedergibt, ist eine Unterscheidung in die Kirchentonarten besser geeignet, Musikstücke zu analysieren und zu vergleichen. Ein Mollstück in Aeolisch-Moll hat einen anderen Charakter als ein Stück in Dorisch-Moll. Als Beispiele für Dorisch-Moll seien hier die Stücke „Drive“ von der Popgruppe „REM“ und „Why does my heart feel so bad“ vom Musikkünstler „Moby“ genannt. Diese unterscheiden sich in ihrem Charakter von „normalen“ Aeolisch-Moll Stücken – wie z.B. das Lied „Horizont“ von Udo Lindenberg oder das Musikstück „Anna“ von der Musikgruppe „Freundeskreis“.

Der neu entwickelte Algorithmus zu diesem Ansatz wird zunächst erklärt und diskutiert. Anschließend erfolgen eine Evaluierung und die Präsentation der Ergebnisse. Der MIREX-Score ist zur Evaluierung ungeeignet, da im ersten Schritt des neuen Ansatzes das richtige Tonset (vgl. Abbildung 11.15) zu erkennen ist. Da es dafür beim MIREX-Score keinen Bewertungsmaßstab gibt, wird in Abschnitt 11.2.2 eine neue Bewertung für die Erkennung von Kirchentonarten eingeführt.

11.2.1 Algorithmus

Der Erkennungsalgorithmus für die Kirchentonarten ist in die zwei Schritte *Tonset-Erkennung* und *Tonika-Bestimmung* unterteilt. Dies ergibt sich aus der Musiktheorie, da auf jedes diatonische Tonset eine der sechs Kirchentonarten gespielt werden kann. Der Algorithmus sucht im richtigen Tonset nach dem am häufigsten gespielten Akkord. Abbildung 11.15 zeigt das dazugehörige Blockschaltbild.

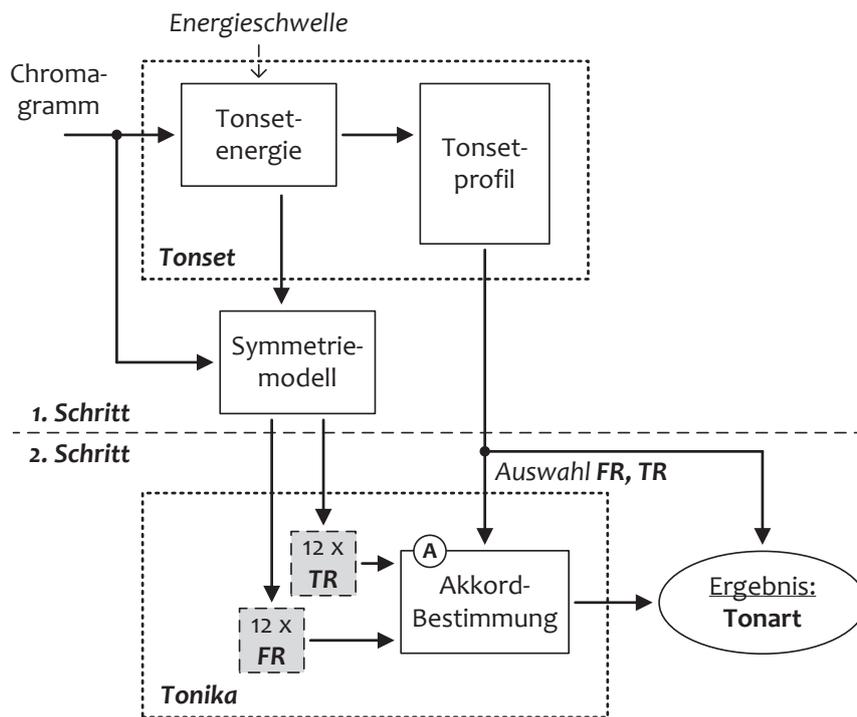


Abbildung 11.15 – Blockschaltbild Erkennung Kirchentonarten

Zur Berechnung des Tonsetprofils und damit zur Bestimmung des richtigen Tonsets wird nur die Tonsetenergie verwendet, da, wie aus Abschnitt 11.1 hervorgeht, die beiden Kirchentonarten Ionisch und Aeolisch durch die Winkelwichtung bevorteilt werden. In dem hier gezeigten Ansatz sind aber alle Kirchentonarten gleichberechtigt. Bei der reinen Dur/Moll-Erkennung hat sich gezeigt, dass die Verwendung der Summenvektorbeträge in den *TR* für die Ermittlung des Tonsetprofils ungeeignet ist. Die Summenvektorbeträge finden daher für die Ermittlung des richtigen Tonsets keine weitere Verwendung. Der Parameter Energieschwelle für die Tonsetbestimmung bei Kirchentonarten wird in Abschnitt 11.2.2 evaluiert. Er hat die gleiche Funktion wie bei der Dur/Moll-Tonarterkennung. In den Kreisen *TR* und *FR* des ermittelten Tonsets wird die Kirchentonartenzuordnung durchgeführt (Bezeichner ‚A‘ in Abbildung 11.15). Abbildung 11.16 zeigt die Zuordnung beispielhaft an den Kreisen des Tonsets von C-Dur, wobei die Grund-

akkorde der Kirchentonarten in allen *TR* und *FR* jeweils die gleiche Lage haben, d.h. Aeolisch und Ionisch befinden sich im *TR* stets oben und Mixolydisch und Dorisch im *FR*. Die Grundakkorde der Kirchentonarten sind im *FR* nebeneinander angeordnet und überschneiden sich im *TR*. Die Überschneidung zeigt Abbildung 11.16 und resultiert aus der Terzschichtung im *TR*. Eine Terz kann in der Regel zwei Akkorden zugeordnet werden. Da der verminderte Akkord hier nicht zu den Grundakkorden gehört, ist in den beiden unteren Terzen des *TR* jeweils nur der Grundakkord einer Kirchentonart zu sehen.

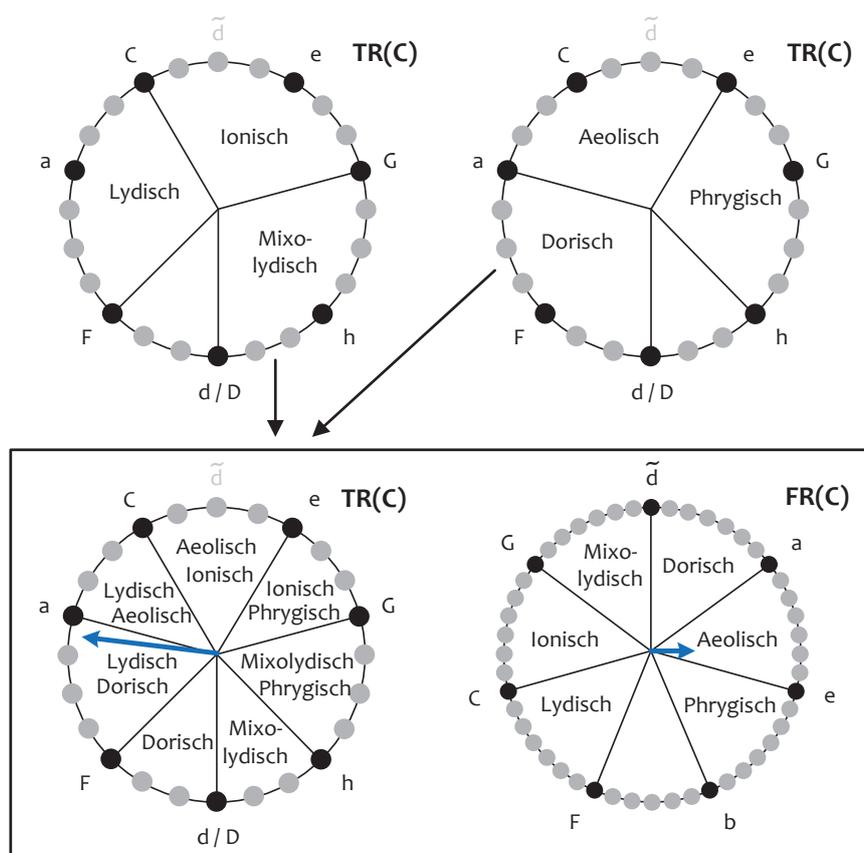


Abbildung 11.16 – Kirchentonarten im TR und FR; Beispiel TR(C), FR(C)

Die Kirchentonartzuordnung geschieht mittels Auszählung der Summenvektoren in den Feldern des *TR* oder *FR*, wobei für jeden gültigen Chromavektor (relative Tonsetenergie oberhalb der Energieschwelle) entweder nur der *TR* oder nur der *FR* verwendet wird. Entscheidend ist der Summenvektorbetrag. Im Beispiel in Abbildung 11.16 ist der Summenvektorbetrag im *TR* größer. Durch die Überschneidung der Tonarten im *TR* werden die Zählwerte der beiden im Trefferfeld enthaltenen Tonarten, hier im Beispiel Lydisch und Dorisch, jeweils um 0,5 erhöht. Im *FR* liegen die sechs Kirchentonarten nebeneinander, und pro Summenvektor in einem dieser Felder wird der entsprechende Zählwert

der Kirchentonart um den Wert 1,0 erhöht. Die Kirchentonart mit dem größten Zählwert wird als Ergebnis ausgewählt. Im nächsten Abschnitt 11.2.2 wird untersucht, ob die Kombination aus *TR* und *FR* für die Grundakkordbestimmung gut geeignet ist oder ob der *TR* bzw. der *FR* jeweils allein bessere Ergebnisse erzielen.

11.2.2 Evaluierung und Ergebnisse

Der neue Algorithmus zur Erkennung von Kirchentonarten wird in diesem Abschnitt evaluiert, und die Ergebnisse werden diskutiert. Dazu werden diejenige *Energieschwelle* und Kombination von *TR* und *FR* ermittelt, bei denen die Tonarterkennungsrate maximal ist. Die Tonarterkennungsrate wird mit einem neuen Bewertungsmaßstab berechnet, der nachfolgend erläutert wird.

11.2.2.1 Bewertungsmaßstab für die Erkennung von Kirchentonarten

Der MIREX-Score ist für die Bewertung von Algorithmen zur Erkennung von Kirchentonarten ungeeignet. Das folgende Beispiel soll dies erklären: Die Tonarten C-Ionisch und C-Mixolydisch sind beides C-Dur Tonarten und würden vom MIREX-Score als gleich angesehen werden, d.h. wenn z.B. die richtige Tonart eines Musikstückes C-Ionisch wäre, der Erkennungsalgorithmus jedoch C-Mixolydisch ausgeben würde, so würde der MIREX-Score das Ergebnis mit 100 % Trefferquote bewerten. Dies ist für die Bewertung der Erkennung von Kirchentonarten allerdings unbrauchbar. Der neue Bewertungsmaßstab bewertet getrennt voneinander, ob das richtige Tonset und ob die richtige Tonika erkannt worden ist.

Tonset	Trefferpunkte	Tonika	Trefferpunkte
Korrektes Tonset	1,0	Korrektter Akkord	1,0
Quintverwandtes Tonset	0,5	Quintverwandter Akkord	0,5
-	-	Terzverwandter Akkord	0,3

Tabelle 11.4 – Bewertungsmaßstab für die Erkennung von Kirchentonarten

Tabelle 11.4 listet dazu die Trefferpunkte auf. Für jedes Stück ergeben sich die Trefferpunkte als Produkt der beiden Werte für *Tonset* und *Tonika*. Die Trefferpunkte sind an den MIREX-Score angelehnt. Für die Tonika – soweit genannt – entsprechen sie denen des MIREX-Scores. Bei der Terzverwandtschaft werden auch die Gegenklänge miteinbezogen

und nicht nur die parallele Tonart wie beim MIREX-Score. Die relativen Tonarten erhalten 0 Trefferpunkte, da sie aus der Sicht der Tonsetverwandtschaft keinen Bezug zu der korrekten Tonart haben. Wenn z.B. anstelle von C-Ionisch die Tonart c-Phrygisch erkannt würde, so wären die Tonsets dieser beiden Tonarten vier Quinten voneinander entfernt (bei einem Maximum von sechs Quinten) und das Tongeschlecht hätte sich von Dur nach Moll geändert. Die Tonarterkennungsrate ergibt sich schließlich, indem die Summe der Trefferpunkte durch die Anzahl der analysierten Musikstücke dividiert wird.

11.2.2.2 Ergebnisse der Evaluierung

In Tabelle 11.5 werden die Ergebnisse der Evaluierung aufgelistet. Die größte Tonarterkennungsrate von insgesamt **68,6 %** wird bei der Verwendung von allen *FR* + *TR* und einer Energieschwelle von 0,89 erzielt. Dabei ist das Ergebnis für die klassischen Stücke mit 74,3 % am größten. Der Grund dafür ist, dass die klassischen Stücke die Tonart sehr klar enthalten und innerhalb der ersten 30 Sekunden so gut wie keine Tonartwechsel vorgenommen werden. Bei der Kategorie *RockPop* gibt es Stücke, z.B. einige vom Künstler Moby, die nur zwei Akkorde enthalten und daher sehr schwierig einer einzigen Tonart zugeordnet werden können. Daher ist hier die Tonarterkennungsrate um 11,4 % niedriger.

	nur <i>FR</i>	nur <i>TR</i>	<i>FR</i> + <i>TR</i>
Energieschwelle	0,92	0,89	0,89
<i>Beatles</i>	57,2 %	64,3 %	68,6 %
<i>RockPop</i>	51,2 %	61,9 %	62,9 %
<i>Klassik</i>	60,8 %	68,2 %	74,3 %
Durchschnitt	56,1 %	64,8 %	68,6 %

Tabelle 11.5 – Ergebnisse der Kirchentonarten-Evaluierung; nach den ersten 30 Sekunden; Parameterevaluierung und Tonarterkennungsraten

Tabelle 11.6 stellt die Ergebnisse getrennt für Tonset und Grundakkord dar. Dabei können zwei Werte einer Zeile nicht einfach multipliziert werden, um auf das jeweilige Ergebnis der Tonarterkennung zu kommen, da es sich um die Mittelwerte der Einzelergebnisse handelt. Es werden jeweils für beide Teilbereiche ca. 80 % richtig erkannt. Es zeigt sich hier, dass vor allem die Tonika bei *RockPop* mit 76,1 % am schlechtesten erkannt wird. Bei *Klassik* sind beide Erkennungsraten mit 87,1 % für das korrekte Tonset und mit 84,0 % für die richtige Tonika am größten.

	Tonset	Tonika
<i>Beatles</i>	81,5 %	79,6 %
<i>RockPop</i>	78,9 %	76,1 %
<i>Klassik</i>	87,1 %	84,0 %
Durchschnitt	82,5 %	79,9 %

Tabelle 11.6 – Ergebnisse der Kirchentonarten-Evaluierung; nach den ersten 30 Sekunden; Erkennungsraten getrennt für Tonset und Tonika; Energieschwelle 0,89

11.3 Zusammenfassung

In diesem Kapitel wurden zwei neue und auf dem Symmetriemodell aufbauende Algorithmen vorgestellt, evaluiert und mit dem Stand der Technik verglichen, wobei der zweite Algorithmus auf einem vollkommen neuen Ansatz zur Tonarterkennung, der ebenfalls hier dargestellt wurde, aufbaut. Der erste Algorithmus zur reinen Erkennung von Dur- und Molltonarten erzielt mit 82,9 % im Durchschnitt das höchste Ergebnis für insgesamt 442 Musikstücke aus drei verschiedenen Musikkategorien und ist somit etwa 20 % besser als der bisherige Stand der Technik. Zum Vergleich wurden andere Algorithmen implementiert und diskutiert. Der zweite Algorithmus ordnet die Tonart eines Musikstückes einer der sechs Kirchentonarten zu. Damit wurde die Grundlage geschaffen, leistungsfähigere Musikbewertungs- und Musiksuchalgorithmen zu entwickeln, da sich die Kirchentonarten besser eignen, den Charakter eines Musikstückes zu beschreiben. Zur Evaluierung der Kirchentonarterkennung wurde zudem ein neues Bewertungsmaß eingeführt. Die Erkennungsrate für die Kirchentonarten beträgt 68,6 % bei den gleichen Teststücken wie für die reine Dur/Moll-Erkennung.

Kapitel 12

Akkorderkennung

Dieses Kapitel beschreibt die Untersuchungen und Ergebnisse zur Akkorderkennung. Im ersten Abschnitt wird erläutert, welche Akkorderkennungsalgorithmen in Kombination mit welchen Merkmalsvektoren evaluiert und verglichen werden. Anschließend werden die Analysealgorithmen evaluiert. Es folgen darauf ein Vergleich der Verfahren und abschließend eine Zusammenfassung der Ergebnisse. Zur Evaluierung der Algorithmen und Merkmalsvektoren werden 100 Beatles-Lieder verwendet. Die Ground Truth mit der symbolischen Beschreibung der ersten zwölf Studioalben der Beatles, in denen die 100 verwendeten Lieder enthalten sind, wurde von Christopher Harte erstellt und ist frei verfügbar in einer Datenbank mit dem Namen „C4DM Chord Toolkit“ (vgl. [57],[1]).

12.1 Gegenstand der Untersuchungen

In diesem Abschnitt werden zu Beginn die eingesetzten Merkmalsvektoren und Analysealgorithmen vorgestellt. Es folgt eine Beschreibung der Ground Truth und der verwendeten Klassen sowie der Trainings- und Analysemodalitäten, bevor die Ergebnisberechnung erläutert wird.

12.1.1 Merkmalsvektoren und Analysealgorithmen

Um zu ermitteln, welcher Merkmalsvektor bei welchem Analysealgorithmus die besten Analyseergebnisse erzielt, werden insgesamt drei Merkmalsvektortypen mit drei Analysealgorithmen kombiniert.

Als Merkmalsvektoren werden verwendet:

- **CHROMA**: Chromavektor. Der Chromavektor entspricht dem Stand der Technik und dient als Vergleichsgrundlage für die anderen beiden Merkmalsvektoren.

- **HSG6D**: 6D-Vektor von Harte, Sandler und Gasser (aus [59]). Der Merkmalsvektor *HSG6D* wurde 2006 im Zusammenhang mit der „Harmonic Change Detection Function“ (vgl. [58]) veröffentlicht und bereits für Tonarterkennung und Akkorderkennung in Verbindung mit HMM von Kyogu Lee und Malcolm Slaney eingesetzt (vgl. [94],[95]), ohne ihn in Beziehung zu dem Chromavektor zu setzen.
- **VSYM**: Merkmalsvektor/Ausgabevektor des in dieser Arbeit neu eingeführten Symmetriemodells. Die verwendeten Kreise und damit Dimensionen werden für die unterschiedlichen Analysealgorithmen einzeln bestimmt. Der vollständige Merkmalsvektor des Symmetriemodells hat 74 Dimensionen (12 *FR*, 12 *TR*, 12 *DR*, 1 *F*; Jeder Kreis enthält zwei Dimensionen).

Als Analysealgorithmen werden verwendet:

- **KNN**: k Nearest Neighbours.
- **HMM**: Hidden Markov Models mit Gaußverteilungen als Wahrscheinlichkeitschätzer. Dieses Maschinenlernverfahren entspricht dem Stand der Technik für die Akkorderkennung und wird in allen neueren Ansätzen verwendet (vgl. Kapitel 9).
- **HMM/KNN**: In dieser Arbeit neu vorgestelltes Verfahren, das anstelle von Gaußverteilungen KNN zur Speicherung der funktionellen Relation verwendet (vgl. Kapitel 8).

Abbildung 12.1 zeigt einen Überblick über das Gesamtanalysesystem.

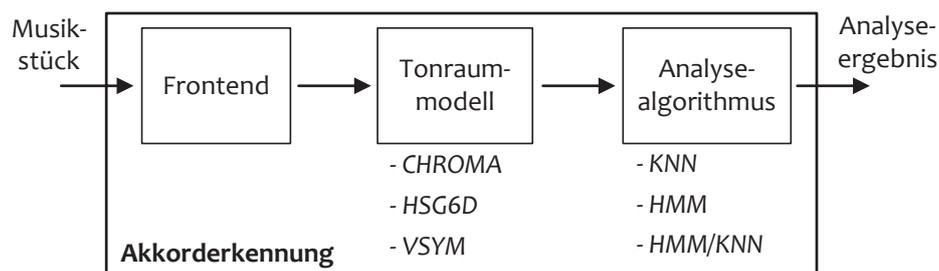


Abbildung 12.1 – Übersicht Tonraummodelle und Analysealgorithmen

12.1.2 Ground Truth und verwendete Klassen

Die Ground Truth von Christopher Harte beschreibt jedes Beatles-Lied einzeln. Sie ist in drei Spalten aufgeteilt, wobei die ersten beiden den Start- und Endzeitpunkt eines Akkords markieren und die dritte den Akkord selbst enthält (vgl. Abbildung 12.2, links).

Ground Truth (Original)			Ground Truth (angepasst):		
Startzeit:	Endzeit:	Akkord:	Startzeit:	Endzeit:	Akkord:
0,000	1,015	N	0,000	1,024	N
1,015	4,155	G:sus4(b7)	1,024	4,992	G
4,155	5,017	G			
5,017	5,853	C	4,992	5,888	C
5,853	7,525	G	5,888	7,552	G
...

(Zeit in Sekunden) (Zeit in Sekunden)

Abbildung 12.2 – Anpassung der Ground Truth an die Testumgebung; „A hard day’s night“ von den Beatles

Die Zeitbasis ist auf eine Millisekunde genau, und als Akkorde werden neben den 24 Dur/Moll-Dreiklängen auch ihre sämtlichen Umkehrungen und zusätzlich alle möglichen Vier- und Fünfklingen verwendet. Eine weitere Klasse ist „N“, die das Ereignis „kein Akkord“ ausdrückt. Dazu gehören Stille, Rauschen und nicht akkordische Tonkombinationen wie z.B. Einzeltöne ohne Kontext und Toncluster. Die Testumgebung wird in dieser Arbeit auf 25 Klassen festgelegt (24 Dur/Moll-Dreiklänge plus die Klasse „N“). Es gibt Ansätze mit einer wesentlich größeren Anzahl unterschiedlicher Klassen; wobei sich gezeigt hat, dass die Mehrdeutigkeit von Drei-, Vier- und Fünfklingen oft zu Klassifizierungsproblemen führt. Als einfaches Beispiel ist ein C-Dur-sixte-ajoutée zu nennen, bei dem die Quinte weggelassen wird. Sobald die Tonigkeiten in Terzschichtung betrachtet werden und der Grundton nicht zusätzlich ausgewertet wird, ist dieser Akkord ein a-Moll-Dreiklang. Zusätzlich wird die Ground Truth mit zunehmender Klassenanzahl und detaillierterer Beschreibung der Akkorde (z.B. C-Dur-Sept-Non-Akkord) immer ungenauer und schwieriger zu erfassen, da die Zeitdauern der einzelnen Klassen immer kürzer werden und die Erstellung der Ground Truth schließlich in anstrengender Kleinarbeit durch den Menschen durchgeführt werden muss. Neben der Anpassung der originalen Klassen an die 25 festgelegten Klassen sind auch die Start- und Endzeitpunkte der originalen Ground Truth so abzuändern, dass sie ein Vielfaches von 128 ms entsprechen, da dies die Zeitrasterung (Schubweite) des Frontends ist. Abbildung 12.2 zeigt dies an einem Ausschnitt der Ground Truth vom Lied „A hard day’s night“ der Beatles. Das Lied beginnt mit 1,015 s Stille, was mit „N“ ausgedrückt wird. Die Endzeit wird auf das nächstliegende Vielfache von 128 ms abgeändert. Dies ist der Zeitpunkt 1024 ms. In der nächsten Zeile wird in der originalen Ground Truth ein Gsus4-Akkord angegeben. Dieser wird auf den Dreiklang G-Dur reduziert. Da die nächste Zeile ebenfalls G-Dur enthält, werden diese beiden Ereignisse zu einem einzigen zusammengefasst (Abbildung 12.2,

rechts, Zeile 2). Die Ground Truth aller verwendeten 100 Beatles-Lieder wird auf diese Art und Weise angepasst, damit sie zum Trainieren der Analysealgorithmen und zur Auswertung der Ergebnisse verwendet werden kann.

12.1.3 Training, Analyse und Berechnung der Ergebnisse

Die 100 verwendeten Beatles Lieder (B1...B100) stammen aus den ersten acht Studioalben der Beatles, wobei aus dem achten Album nur die ersten drei Lieder verwendet werden, um insgesamt auf 100 Teststücke zu kommen. Die Analyseergebnisse beziehen sich stets auf ein ganzes Album. Die Einzelergebnisse dazu sind in Anhang F zu finden. Um objektive Ergebnisse zu erhalten, unterscheiden sich die Trainingsstücke immer von den Teststücken (Kreuzvalidierung), d.h., wenn z.B. das erste Beatles-Album „Please Please Me“ (B1...B14) untersucht wird, so werden als Trainingsstücke die verbleibenden Lieder (B15...B100) verwendet.

Ground Truth (angepasst):			Akkordliste der Analyse:	
Startzeit:	Endzeit:	Akkord:		Akkord:
0,000	0,128	N		N
0,128	0,256	N		N
0,256	0,384	N		N
0,384	0,512	N		N
0,512	0,640	N		N
0,640	0,768	N		N
0,768	0,896	N		N
0,896	1,024	N		N
1,024	1,152	G		N
1,152	1,280	G		G
1,280	1,408	G		G
...

(Zeit in Sekunden)

Vergleich der Ground Truth und Akkordliste

Falsch erkannter Akkord

Abbildung 12.3 – Berechnung der Trefferquote; „A hard day’s night“ von den Beatles

Die Analyse eines Musikstückes liefert eine Liste an Akkorden (Klassen). Dies stellt Abbildung 12.3 dar. Auf der linken Seite ist die angepasste Ground Truth aus Abbildung 12.2 zu sehen und rechts daneben die Akkordliste der Analyse. Ein Vergleich der beiden Listen liefert das Analyseergebnis (Trefferquote), indem die Anzahl der Zeitfenster mit korrekt erkannten Akkorden durch die Gesamtanzahl der Zeitfenster geteilt wird. Für das dargestellte Beispiel beträgt die Trefferquote $10/11 = 90,9\%$.

12.2 Akkorderkennung mit KNN

Dieser Abschnitt betrachtet das Maschinenlernverfahren KNN in Verbindung mit den drei vorgestellten Merkmalsvektoren. Die zu evaluierenden Parameter bei KNN sind die „Anzahl der Trainingspunkte pro Klasse“ (TP) während des Trainings, die „Anzahl der zu suchenden Nachbarn“ (K) während der Analyse sowie die Wahl des Abstandsmaßes.¹ Wie in Abschnitt 8.2.1 beschrieben wird generell das *gewichtete Auszählen* verwendet. Zuerst werden der Merkmalsvektoren CHROMA und HSG6D evaluiert. Anschließend wird der Merkmalsvektor VSYM des Symmetriemodells untersucht und die Ergebnisse miteinander verglichen.

12.2.1 Evaluierung CHROMA

Zur Evaluierung des Merkmalsvektors CHROMA wird zu Beginn bestimmt, welches Abstandsmaß die besten Ergebnisse erzielt. Dafür werden Werte für TP und K angenommen. Es werden $TP = 100$ Trainingspunkte pro Klasse und $K = 1500$ nächste Nachbarn für die Analyse vorgegeben. Zur sinnvollen Darstellung der Evaluierungsergebnisse wird in den nachfolgenden Tabellen anstelle von K das Verhältnis K/TP angegeben. Für $K = 1500$ und $TP = 100$ hat K/TP den Wert 15. Das Maximum für K/TP ist 25, da es insgesamt 25 Klassen gibt. Tabelle 12.1 zeigt das Ergebnis der Abstandsevaluierung. Die Manhattan-Distanz erzielt mit 48,2 % für das erste Beatles-Album „Please Please Me“ das beste Ergebnis.

Kosinus	Manhattan	Euklidisch
45,4	48,2	45,6

Tabelle 12.1 – KNN+CHROMA; Evaluierung Abstandsmaß; Trefferquote in Prozent; $TP = 100$; $K/TP = 15$; Training=B15...B100; Test=B1...B14 (Beatles-Album „Please Please Me“)

Zur Ermittlung von TP und K wird eine Analysereihe durchgeführt mit unterschiedlichen Werten für TP und K/TP . Tabelle 12.2 listet die Ergebnisse der Evaluierung auf. Das beste Ergebnis (dunkelgrau hinterlegt) wird für $TP = 500$ und $K/TP = 3$ erzielt. Die hellgrau hinterlegten Felder geben den Ergebnisbereich an, der innerhalb von 5% unter dem Maximum liegt. Je größer dieser Bereich ist, desto stabiler und sicherer ist das Ergebnis.

¹ Vgl. Abschnitt 8.2.1.

		TP - Anzahl Trainingspunkte pro Klasse						
K/TP	1	5	10	50	100	200	300	500
1	39,6	44,3	45,2	47,5	48,8	50,4	51,4	52,2
3	39,6	41,3	42,9	46,6	48,2	50,1	51,0	52,3
10	39,6	43,1	44,2	46,6	47,7	49,4	50,6	52,1
25	39,6	44,0	45,0	47,4	48,5	50,1	50,8	52,1

Tabelle 12.2 – KNN+CHROMA; Evaluierung TP und K/TP ; Manhattan-Distanz; Trefferquote in Prozent; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)

12.2.2 Evaluierung HSG6D

Der Merkmalsvektor HSG6D von Harte, Sandler und Gasser wird auf die gleiche Weise evaluiert wie der Merkmalsvektor CHROMA. Hierbei liefert auch die Manhattan-Distanz die größte Trefferquote (siehe Tabelle 12.3).

Kosinus	Manhattan	Euklidisch
40,2	40,5	40,3

Tabelle 12.3 – KNN+HSG6D; Evaluierung Abstandsmaß; Trefferquote in Prozent; $TP = 100$; $K/TP = 15$; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)

Nach der Bestimmung des optimalen Abstandsmaßes werden TP und K ermittelt. Für $TP = 500$ und $K/TP = 3$ ist die Trefferquote am größten (siehe Tabelle 12.4).

		TP - Anzahl Trainingspunkte pro Klasse						
K/TP	1	5	10	50	100	200	300	500
1	34,8	36,5	37,2	38,5	40,0	41,5	42,1	42,8
3	34,9	36,7	37,2	39,5	40,9	42,5	43,2	44,1
10	34,9	36,5	36,8	38,9	40,5	42,3	43,4	44,0
25	34,8	36,5	36,8	38,8	40,4	42,1	43,0	43,8

Tabelle 12.4 – KNN+HSG6D; Evaluierung TP und K/TP ; Manhattan-Distanz; Trefferquote in Prozent; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)

Die Ergebnisse für das erste Beatles-Album sind ca. 8 % niedriger als bei Verwendung des Merkmalsvektors CHROMA. In Abschnitt 12.2.4 wird gezeigt, dass der Vergleich mit einem Album (14 Teststücke) nicht ausreicht, um eine sichere Aussage über die Leistungsfähigkeit der verschiedenen Analysealgorithmen treffen zu können.

12.2.3 Evaluierung VSYM

Vor der eigentlichen Evaluierung wird der Merkmalsvektor VSYM in Bezug auf harmonische Ähnlichkeit betrachtet. Kapitel 10 kommt zum Ergebnis, dass der Winkel der Summenvektoren in den Symmetriemodell-Kreisen die harmonische Information trägt und die Länge der Summenvektoren hierbei keine Funktion übernimmt. Das Diagramm (a) in Abbildung 12.4 zeigt als Beispiel die gewöhnliche Ausgabe eines Kreises (z.B. *TR*) im Symmetriemodell. Der blaue Vektor ist ein gespeicherter Trainingsvektor. Es ist nun zu ermitteln, ob der grüne oder rote Analysevektor harmonisch näher am Trainingsvektor ist. Der Abbildung ist zu entnehmen, dass der Abstand (d_2) vom roten Vektor geringer ist. Dies liegt daran, dass der blaue Trainingsvektor und der rote Testvektor jeweils einen kleinen Betrag haben. Der Winkelabstand vom blauen Trainingsvektor zum grünen Testvektor ist jedoch wesentlich geringer als zum roten Testvektor. Daher ist auch der grüne Testvektor harmonisch näher am Trainingsvektor als der rote. Die Vektoren werden auf die Länge 1 normiert, damit der geometrische Abstand dem harmonischen Abstand entspricht. Dies ist im Diagramm (b) in Abbildung 12.4 zu sehen.

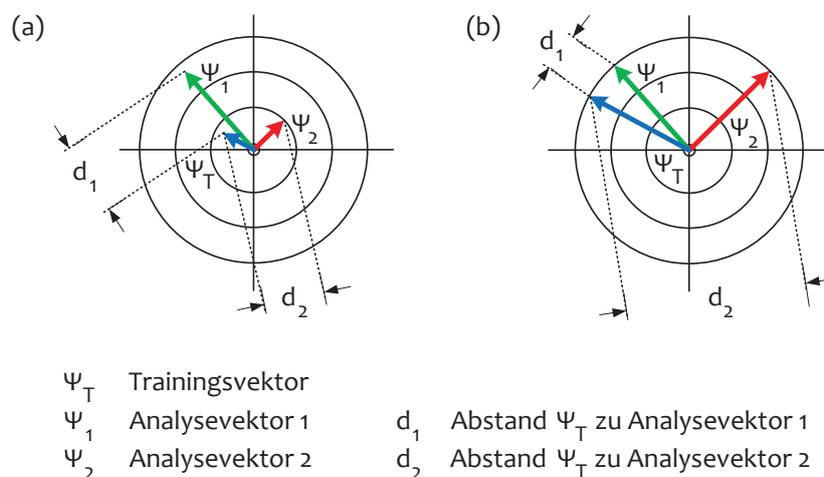


Abbildung 12.4 – Symmetriemodell; (a) nicht normierte Vektoren; (b) normierte Vektoren

Der vollständige Ausgabevektor des Symmetriemodells (ohne die 12 *CR*) besteht aus 74 Dimensionen. In Tabelle 12.5 werden verschiedene Dimensionen des Symmetriemodells ausgewählt und für die drei verwendeten Abstandsmaße verglichen. *TR*, *FR* und *DR* stehen dabei jeweils für alle 12 dieser tonartbezogenen Kreise. Die *TR*-Kreise und *FR*-Kreise ergeben einzeln die höchsten Trefferquoten. Dies ist ein Hinweis darauf, dass für harmonische Analysen diese Kreise am besten geeignet sind. Das insgesamt beste Analyseergebnis ergibt sich für die Kombination aus *TR*, *FR* und *DR* in Verbindung mit der euklidischen Distanz.

	Kosinus	Manhattan	Euklidisch
TR	39,5	35,0	37,8
FR	40,0	39,5	39,4
DR	32,1	29,2	30,2
F	15,3	13,0	10,5
TR+FR	48,7	48,3	49,0
TR+FR+DR	49,6	48,8	50,4
TR+FR+DR+F	49,5	48,8	50,4

Tabelle 12.5 – KNN+VSYM; Evaluierung Abstandsmaß und Symmetriemodell-Kreise; Vektoren normiert; Trefferquote in Prozent; $TP = 100$; $K/TP = 15$; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)

In Tabelle 12.6 und Tabelle 12.7 werden die Ergebnisse für die Evaluierung der Werte TP und K dargestellt. Tabelle 12.6 zeigt die Ergebnisse für den nicht normierten Merkmalsvektor VSYM und Tabelle 12.7 für den normierten Merkmalsvektor. Die Ergebnisse verbessern sich bei Normierung des Merkmalsvektors um gut 10 % von 43,2 % auf 53,5 %. Dies ergibt sich bei $TP = 500$ und $K/TP = 3$.

TP - Anzahl Trainingspunkte pro Klasse								
K/TP	1	5	10	50	100	200	300	500
1	40,1	42,1	41,1	41,9	41,8	41,9	42,0	42,3
3	40,1	41,8	42,2	42,8	43,0	43,2	43,0	42,8
10	40,1	41,4	41,8	42,4	42,6	42,4	43,2	43,1
25	40,1	41,4	41,8	42,2	42,3	42,5	42,8	42,8

Tabelle 12.6 – KNN+VSYM(FR+TR+DR); Vektoren nicht normiert; Evaluierung TP und K/TP ; Euklidische Distanz; Trefferquote in Prozent; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)

12.2.4 Ergebnisse und Vergleich

In Tabelle 12.8 werden die Ergebnisse der ersten sieben Beatles-Alben in Kombination mit den drei Merkmalsvektoren aufgelistet. Es werden zur Ergebnisberechnung jeweils die ermittelten Parameter für TP und K verwendet. Der Merkmalsvektor VSYM hat mit 51,1 % im Durchschnitt die höchste Trefferquote im Vergleich zum Merkmalsvektor CHROMA mit 46,2 % und zum Merkmalsvektor HSG6D mit 44,3 %. Obwohl sich CHROMA und HSG6D beim ersten Beatles-Album um über 8 % unterscheiden, differieren die Mittelwerte für die sieben Beatles-Alben um nur knapp 2 %. Dies zeigt, dass 14 Teststücke (B1... B14) nicht ausreichend sind, um Algorithmen allgemeingültig zu vergleichen.

		TP - Anzahl Trainingspunkte pro Klasse						
K/TP	1	5	10	50	100	200	300	500
1	46,2	46,6	46,9	47,9	49,1	50,3	51,4	52,4
3	46,2	47,8	47,9	49,2	50,5	51,8	52,5	53,5
10	46,2	47,2	47,4	48,8	50,2	51,7	52,8	53,5
25	46,2	47,1	47,3	48,6	50	51,5	52,5	53,3

Tabelle 12.7 – KNN+VSYM(FR+TR+DR); Vektoren normiert; Evaluierung *TP* und *K/TP*; Treffer in Prozent; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)

Beatles-Album	Teststücke	CHROMA	HSG6D	VSYM
Please Please Me	B1...B14	52,3	44,0	53,5
With The Beatles	B15...B28	45,3	39,4	44,5
A Hard Day's Night	B29...B41	50,3	44,6	48,8
Beatles For Sale	B42...B55	58,6	52,2	60,8
Help	B56...B69	40,5	48,9	55,3
Rubber Soul	B70...B83	36,8	43,5	49,2
Revolver	B84...B97	39,6	37,8	45,3
	Mittelwert	46,2	44,3	51,1

Tabelle 12.8 – Ergebnisse KNN; Trefferquote in Prozent; $TP = 500$; $K/TP = 3$;

Der Test auf Signifikanz wird mit dem Friedman-Test anhand der Ergebnisliste aller 100 Teststücke durchgeführt. Es ergibt sich, dass mit einer Fehlerwahrscheinlichkeit von $\alpha = 5\%$ die Mittelwerte aller drei Messreihen und damit auch die Verwendung der Merkmalsvektoren für das Analyseverfahren KNN signifikant unterschiedlich sind (siehe Abbildung 12.5).

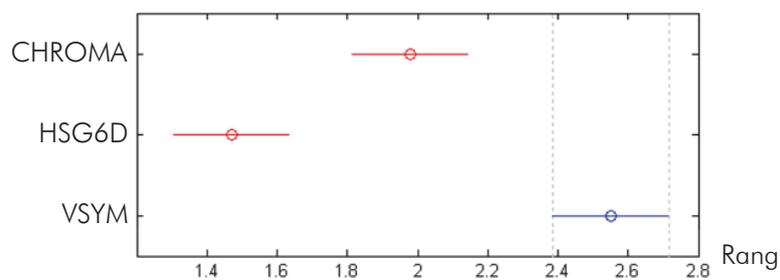


Abbildung 12.5 – Ergebnisse KNN; Friedman-Test; $\alpha = 5\%$

12.3 Akkorderkennung mit HMM

In diesem Abschnitt wird untersucht, welcher Merkmalsvektor in Verbindung mit dem Maschinenlernverfahren HMM die besten Ergebnisse erzielt. Es wird eine mehrdimensionale Gaußverteilung als Wahrscheinlichkeitsschätzer benutzt. Dies entspricht dem Stand der Technik. Das HMM $\lambda = \{A, b, \pi\}$ wird – wie in Abschnitt 8.2.2 beschrieben – trainiert und kann auf verschiedene Arten evaluiert werden. Diese werden bereits von Bello und Pickens in [12] sowie Papadopoulos und Peeters in [121] ausführlich diskutiert und hier nicht weiter aufgegriffen. In dieser Arbeit wird zur Schätzung der Wahrscheinlichkeiten bei der Analyse die Diagonale der einzelnen Kovarianzmatrizen verwendet und die restlichen Kovarianzen zu Null gesetzt. Darüber hinaus sind keine weiteren Parameter des HMM zu evaluieren. Es folgen die Evaluierung der Dimensionen (Kreise) des Merkmalsvektors VSYM und anschließend die Präsentation und der Vergleich der Ergebnisse.

12.3.1 Evaluierung VSYM

Bei KNN hat sich gezeigt, dass die Kombinationen *TR+FR* und *TR+FR+DR* am besten für die Akkordanalyse geeignet sind. Die Evaluierung bei HMM ergibt, dass die Kombination *TR+FR* die höchste Trefferquote erzielt (Vgl. Tabelle 12.9).

TR+FR	TR+FR+DR
56,5	55,4

Tabelle 12.9 – HMM+VSYM; Evaluierung Symmetriemodell-Kreise; Trefferquote in Prozent; Training=B15. . . B100; Test=B1. . . B14 (Beatles-Album „Please Please Me“)

12.3.2 Ergebnisse und Vergleich

Tabelle 12.10 stellt die Ergebnisse der Akkordanalyse für HMM und die drei unterschiedlichen Merkmalsvektoren dar. Der Merkmalsvektor CHROMA führt zur geringsten Trefferquote mit durchschnittlich 49,9 % bei den sieben Beatles-Alben. Der Merkmalsvektor VSYM hat mit 54,3 % die zweithöchste und der Merkmalsvektor HSG6D mit 62,1 % die höchste Trefferquote. Der Test auf Signifikanz ergibt, dass der Merkmalsvektor HSG6D mit einer Fehlerwahrscheinlichkeit von $\alpha = 5\%$ bei HMM für Akkorderkennung signifikant besser geeignet ist als die anderen beiden Merkmalsvektoren (Vgl. Abbildung 12.6). Das Ergebnis lässt schließen, dass die Verteilung der Klassen (Akkorde) im HSG6D-Merkmalsraum gaußähnlich und daher gut für das verwendete HMM geeignet ist. Im Gegenzug ist es sehr wahrscheinlich, dass die Klassen im CHROMA- und

Beatles-Album	Teststücke	CHROMA	HSG6D	VSYM
Please Please Me	B1...B14	56,6	61,9	55,8
With The Beatles	B15...B28	51,5	56,9	44,6
A Hard Day's Night	B29...B41	58,8	56,2	53,1
Beatles For Sale	B42...B55	62,5	77,6	64,8
Help	B56...B69	37,5	67,7	57,3
Rubber Soul	B70...B83	39,7	63,0	55,0
Revolver	B84...B97	42,8	51,3	49,6
	Mittelwert	49,9	62,1	54,3

Tabelle 12.10 – Ergebnisse HMM; Trefferquote in Prozent

VSYM-Merkmalsraum nicht gaußähnlich verteilt sind und durch das Abbilden auf eine Gaußverteilung Fehler entstehen, die eine niedrigere Trefferquote in der Analyse als Folge haben.

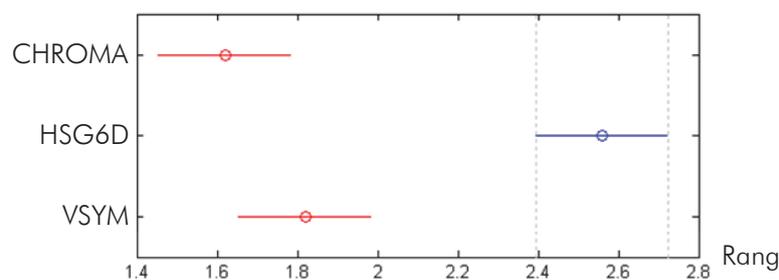


Abbildung 12.6 – Ergebnisse HMM; Friedman-Test; $\alpha = 5\%$

12.4 Akkorderkennung mit HMM/KNN

Dieser Abschnitt untersucht das hier neu vorgestellte Analyseverfahren, das HMM und KNN kombiniert. Bei diesem Verfahren wird der Wahrscheinlichkeitsschätzer, der Gaußverteilungen verwendet, durch einen KNN-Wahrscheinlichkeitsschätzer ersetzt. Das neue Analyseverfahren wird wie die anderen beiden schon betrachteten Verfahren mit den drei unterschiedlichen Merkmalsvektoren evaluiert und die Ergebnisse werden anschließend verglichen. Die Evaluierung entspricht der des Analyseverfahrens KNN, weil das Analyseverfahren KNN komplett im Verfahren HMM/KKN enthalten ist. Daher sind die Parameter TP , K und das Abstandsmaß zu bestimmen sowie das *gewichtete Auszählen* zu verwenden.

12.4.1 Evaluierung CHROMA

Die Kosinus-Distanz ergibt bei den vorgegebenen Parametern $TP = 100$ und $K/TP = 15$ die höchste Trefferquote und wird als Abstandsmaß ausgewählt (Vgl. Tabelle 12.11).

Kosinus	Manhattan	Euklidisch
72,4	56,9	68,5

Tabelle 12.11 – HMM/KNN+CHROMA; Evaluierung Abstandsmaß; Trefferquote in Prozent; $TP = 100$; $K/TP = 15$; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)

Die Evaluierung der Parameter TP und K zeigt, dass bereits bei wenigen Trainingspunkten pro Zustand $TP = 30$ die höchste Trefferquote mit 73,0 % für das erste Beatles-Album „Please Please Me“ erreicht wird (Vgl. Tabelle 12.12).

K/TP	TP - Anzahl Trainingspunkte pro Klasse							
	1	5	10	30	100	200	300	500
1	41,6	49,5	52,6	56,3	60,1	62,6	64,1	66,5
3	52,8	58,5	60,8	63,6	66,9	68,5	69,2	70,7
10	61,2	65,7	69,7	72,3	72,6	71,9	71,6	71,8
15	66,6	68,9	70,4	73,0	72,4	72,3	71,8	71,2
25	68,9	69,4	70,8	72,9	72,8	72,5	71,6	72,1

Tabelle 12.12 – HMM/KNN+CHROMA; Evaluierung TP und K/TP ; Kosinus-Distanz; Trefferquote in Prozent; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)

Die Anzahl der zu suchenden Nachbarn ist hier $K = 450$. Das Ergebnis ist über einen weiten Wertebereich stabil (grau hinterlegter Bereich). Bei $TP = 10$ und $K = 100$ beträgt die Trefferquote schon 69,7 %. Bei einem einzigen Trainingspunkt pro Zustand ($TP = 1$) und 15 nächsten Nachbarn ($K = 15$) wird eine Trefferquote von 66,6 % erreicht. Dieser Wert liegt bereits 4 % über dem besten Ergebnis des Verfahrens HMM.

12.4.2 Evaluierung HSG6D

Die euklidische Distanz ergibt beim Merkmalsvektor HSG6D die höchste Trefferquote (Vgl. Tabelle 12.13).

In Tabelle 12.14 sind die Ergebnisse der Evaluierung der Parameter TP und K aufgelistet. Das Maximum beträgt 67,6 % (bei $TP = 300$ und $K = 4500$). Im Gegensatz zum Merkmalsvektor CHROMA nimmt die Trefferquote bei der Variation von TP schneller ab und ist daher nicht so stabil.

Kosinus	Manhattan	Euklidisch
64,3	60,2	65,5

Tabelle 12.13 – HMM/KNN+HSG6D; Evaluierung Abstandsmaß; Trefferquote in Prozent; $TP = 100$; $K/TP = 15$; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)

K/TP	TP - Anzahl Trainingspunkte pro Klasse							
	1	5	10	50	100	200	300	500
1	38,7	47,0	49,1	53,9	55,9	57,8	59,3	61,3
3	49,9	55,8	57,4	59,2	60,4	62,3	64,3	65,7
10	56,8	59,3	60,8	62,7	64,1	65,6	66,7	68,5
15	61,0	62,6	63,3	64,3	65,5	66,7	67,6	68,7
25	63,8	64,5	64,7	65,5	65,9	67,1	67,5	68,8

Tabelle 12.14 – HMM/KNN+HSG6D; Evaluierung TP und K/TP ; Euklidische Distanz; Trefferquote in Prozent; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)

12.4.3 Evaluierung VSYM

Zur Evaluierung des Merkmalsvektors VSYM werden verschiedene Dimensionen (Kreise des Symmetriemodells) ausgewählt und bei den Abstandsmaßen Kosinus-Distanz und euklidische Distanz verglichen. Die Manhattan-Distanz hat bei KNN die geringsten Trefferquoten ergeben und wird daher nicht in den Vergleich miteinbezogen. Die Kombinationen $TR+FR$, $TR+FR+DR$ und $TR+FR+DR+F$ ergeben sehr ähnliche Trefferquoten (Vgl. Tabelle 12.15). Von diesen wird die Kombination mit den wenigsten Dimensionen ausgewählt ($TR+FR$), da hierbei der Rechenaufwand am geringsten ist.

	Kosinus	Euklidisch
TR	70,1	70,7
FR	62,4	61,5
DR	57,5	52,9
F	23,3	18,8
TR+FR	76,5	75,8
TR+FR+DR	76,6	74,4
TR+FR+DR+F	76,5	74,5

Tabelle 12.15 – HMM/KNN+VSYM; Evaluierung Abstandsmaß und Symmetriemodell-Kreise; Trefferquote in Prozent; $TP = 100$; $K/TP = 15$; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)

Tabelle 12.16 zeigt die Ergebnisse der Evaluierung von TP und K . Für $TP = 100$ und $K = 1500$ wird eine maximale Trefferquote von 76,5 % erreicht. Das Ergebnis ist über einen weiten Bereich sehr stabil. Bei nur einem Trainingspunkt pro Zustand ($TP = 1$) und 15 nächsten Nachbarn ($K = 15$) beträgt die Trefferquote bereits 72,8 %.

K/TP	TP - Anzahl Trainingspunkte pro Klasse							
	1	5	10	50	100	200	300	500
1	46,1	52,0	54,5	60,0	62,0	64,5	64,4	67,3
3	55,7	60,7	62,3	66,7	68,7	70,9	71,2	71,8
10	69,0	70,8	72,4	74,1	75,2	75,6	76,0	75,2
15	72,8	73,9	74,2	75,6	76,5	76,2	76,4	75,4
25	74,6	75,2	75,3	75,7	76,3	76,2	76,3	76,2

Tabelle 12.16 – HMM/KNN+VSYM(FR+TR); Vektoren normiert; Evaluierung TP und K/TP ; Kosinus-Distanz; Trefferquote in Prozent; Training=B15...B100; Test=B1...B14 (Beatles-Album „Please Please Me“)

12.4.4 Ergebnisse und Vergleich

Die Ergebnisse für das neue Analyseverfahren HMM/KNN kombiniert mit den drei verschiedenen Merkmalsvektoren sind in Tabelle 12.17 zu sehen. Die Merkmalsvektoren CHROMA und VSYM sind für das Maschinenlernverfahren HMM gleichwertig und erzielen die besten Ergebnisse. Der Merkmalsvektor CHROMA erzielt durchschnittlich 72,5 % und der Merkmalsvektor VSYM 72,8 % für die sieben Beatles-Alben, der Merkmalsvektor HSG6D nur 67,8 %.

Beatles-Album	Teststücke	CHROMA	HSG6D	VSYM
Please Please Me	B1...B14	73,0	67,6	76,2
With The Beatles	B15...B28	68,1	60,3	66,1
A Hard Day's Night	B29...B41	72,8	66,1	72,8
Beatles For Sale	B42...B55	83,0	80,7	83,7
Help	B56...B69	77,2	74,8	79,0
Rubber Soul	B70...B83	68,9	66,2	67,6
Revolver	B84...B97	64,5	58,6	64,1
	Mittelwert	72,5	67,8	72,8

Tabelle 12.17 – Ergebnisse HMM/KNN; Trefferquote in Prozent

Der Friedman-Test zeigt, dass die beiden Merkmalsvektoren CHROMA und VSYM in Kombination mit dem Verfahren HMM/KNN mit einer Fehlerwahrscheinlichkeit von $\alpha = 5\%$ signifikant besser sind als der Merkmalsvektor HSG6D (Vgl. Abbildung 12.7).

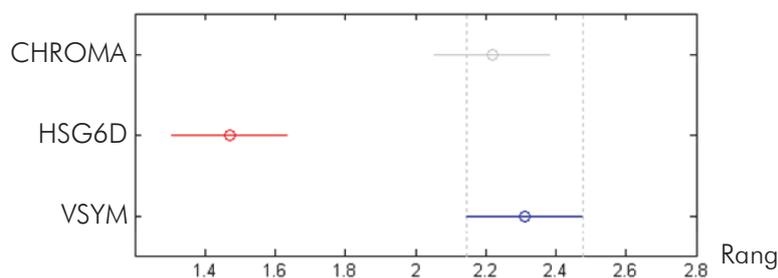


Abbildung 12.7 – Ergebnisse HMM/KNN; Friedman-Test; $\alpha = 5\%$

12.5 Vergleich der Merkmalsvektoren und Algorithmen

Abbildung 12.8 zeigt als Überblick und statistische Zusammenfassung der letzten drei Abschnitte das Ergebnis des Friedman-Tests aller neun möglichen Kombinationen aus den drei Merkmalsvektoren und den drei verwendeten Analysealgorithmen.

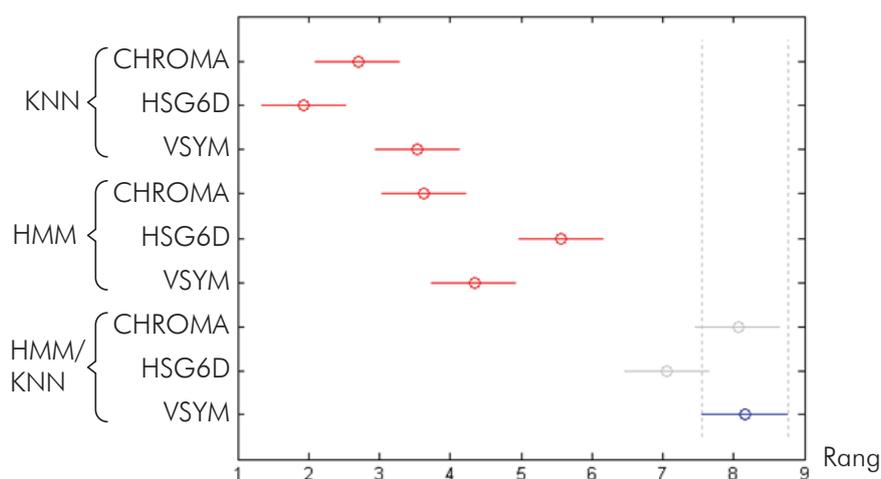


Abbildung 12.8 – Ergebnisse Akkorderkennung; Überblick; Friedman-Test; $\alpha = 5\%$

12.5.1 Vergleich der Merkmalsvektoren

Es werden jeweils die Ergebnisse der Verfahren zusammengefasst, die mit denselben Merkmalsvektoren arbeiten. Die Werte für den Merkmalsvektor CHROMA ergeben sich z.B. aus dem Durchschnitt der Verfahren KNN+CHROMA, HMM+CHROMA und HMM/KNN+CHROMA. Tabelle 12.18 listet die Ergebnisse der drei Merkmalsvektoren CHROMA, HSG6D und VSYM im Durchschnitt auf. Der Merkmalsvektor VSYM erreicht im Mittel 59,4 % und ist damit der Merkmalsvektor mit der höchsten Trefferquote. Der Merkmalsvektor HSG6D kommt auf 58,1 % und der Merkmalsvektor CHROMA auf 56,2 %.

Beatles-Album	Teststücke	CHROMA	HSG6D	VSYM
Please Please Me	B1...B14	60,6	57,8	61,8
With The Beatles	B15...B28	55,0	52,2	51,7
A Hard Day's Night	B29...B41	60,7	55,7	58,2
Beatles For Sale	B42...B55	68,0	70,2	69,8
Help	B56...B69	51,7	63,8	63,8
Rubber Soul	B70...B83	48,5	57,6	57,3
Revolver	B84...B97	48,9	49,2	53,0
	Mittelwert	56,2	58,1	59,4

Tabelle 12.18 – Ergebnisse Akkorderkennung im Durchschnitt; **Merkmalsvektoren**; Trefferquote in Prozent

Abbildung 12.9 stellt die statistische Auswertung hierzu dar. Die Merkmalsvektoren VSYM und HSG6D unterscheiden sich nicht signifikant. Gegenüber dem Stand der Technik (Merkmalsvektor CHROMA) ist der neue Merkmalsvektor VSYM bei den verwendeten Analysealgorithmen jedoch signifikant besser.

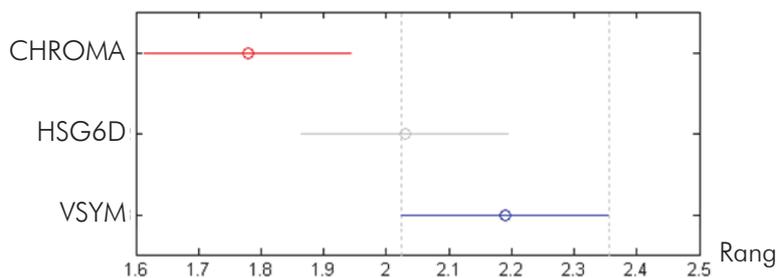


Abbildung 12.9 – Ergebnisse Akkorderkennung; **Vektoren**; Friedman-Test; $\alpha = 5\%$

12.5.2 Vergleich der Analysealgorithmen

In Tabelle 12.19 werden die drei Analysealgorithmen KNN, HMM und HMM/KNN gegenübergestellt. Hierfür werden jeweils die Ergebnisse zusammengefasst, die mit demselben Maschinenlernverfahren erzeugt worden sind. Das neue Verfahren HMM/KNN erreicht mit 71,0 % im Mittel die höchste Trefferquote und damit eine um fast 16 % höhere Trefferquote als das gewöhnliche HMM, das eine Gaußverteilung zum Schätzen der Wahrscheinlichkeiten verwendet. Dies zeigt, dass der KNN-Wahrscheinlichkeitsschätzer in Kombination mit HMM besser als Gaußverteilungen für Akkorderkennung geeignet ist. Das Verfahren KNN erreicht mit im Mittel 47,2 % die geringste Trefferquote. Dies war zu erwarten, da KNN nur eine funktionelle und nicht wie die anderen beiden Verfahren zusätzlich eine zeitliche Relation verwendet.

Beatles-Album	Teststücke	KNN	HMM	HMM/KNN
Please Please Me	B1...B14	49,9	58,1	72,2
With The Beatles	B15...B28	43,1	51,0	64,8
A Hard Day's Night	B29...B41	47,9	56,1	70,6
Beatles For Sale	B42...B55	57,2	68,3	82,5
Help	B56...B69	48,2	54,2	77,0
Rubber Soul	B70...B83	43,2	52,5	67,6
Revolver	B84...B97	40,9	47,9	62,4
	Mittelwert	47,2	55,4	71,0

Tabelle 12.19 – Ergebnisse Akkorderkennung im Durchschnitt; **Analysealgorithmen**; Trefferquote in Prozent

Abbildung 12.10 zeigt das Ergebnis der statistischen Auswertung. Als Fehlerschranke wird $\alpha = 0,3\%$ verwendet. Das neue Verfahren HMM/KNN ist hoch signifikant besser als die anderen beiden Verfahren KNN und HMM. Die Signifikanz geht bereits aus Abbildung 12.8 hervor, da hier die Kombinationen aus HMM/KNN und den drei unterschiedlichen Merkmalsvektoren CHROMA, HSG6D und VSYM bereits signifikant vor den anderen Verfahren liegen.

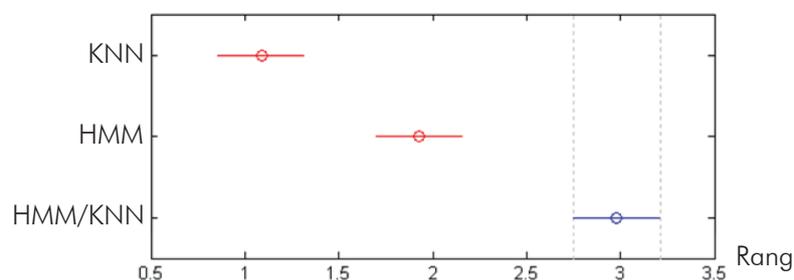


Abbildung 12.10 – Ergebnisse Akkorderkennung; **Algorithmen**; Friedman-Test; $\alpha = 0,3\%$

12.6 Vergleich mit dem Stand der Technik

Die Ergebnisse des neuen HMM/KNN-Algorithmus' werden mit dem Stand der Technik verglichen. Dafür werden die beiden Beatles-Alben „Please Please Me“ (CD1) und „Beatles for Sale“ (CD2) verwendet.² Das Training erfolgt jeweils mit den anderen Beatles-Alben. Tabelle 12.20 listet die Ergebnisse der verschiedenen Algorithmen auf. Eine genauere Beschreibung der einzelnen Algorithmen ist in Kapitel 9 zu finden.

² Vgl. [60], [12], [94], [157] und Abschnitt 9.2.

Autor	Algorithmus	CD1	CD2	Mittelwert
Harte/Sandler	Korrelation mit Chromavektoren	53,9	70,8	62,4
Bello/Pickens	HMM	68,6	81,6	75,1
Lee/Slaney	HMM	64,0	84,7	74,4
Sumi und Kollegen	HMM + Basserkennung + Rhythmuserkennung	75,1	79,7	77,4
(diese Arbeit)	HMM	61,9	77,6	69,8
(diese Arbeit)	HMM/KNN	76,2	83,7	80,0

Tabelle 12.20 – Ergebnisse Akkorderkennung; Vergleich mit Stand der Technik: Harte/Sandler [56], Bello/Pickens [12], Lee/Slaney [94] und Sumi und Kollegen [157]; Trefferquote in Prozent

Der Algorithmus von Harte/Sandler nutzt einen einfachen Korrelations-Algorithmus und erreicht deshalb mit 62,4 % die geringste Trefferquote. Die beiden Ansätze von Bello/Pickens und Lee/Slaney, welche HMM verwenden, kommen auf 75,1 % bzw. 74,4 %. Diese Ergebnisse werden in dieser Arbeit mit dem gleichen Ansatz nicht ganz erzielt, wobei zu erwähnen ist, dass im Stand der Technik die Trainings- und Teststücke identisch sind, was in der Regel zu besseren Ergebnissen führt. Der hier verwendete HMM-Algorithmus erzielt in Verbindung mit dem Merkmalsvektor HSG6D eine Trefferquote von durchschnittlich 69,8 % für die beiden betrachteten Beatles-Alben. Der bisher beste und komplexeste Ansatz wurde 2008 von Sumi und Kollegen bei der ISMIR2008³ vorgestellt. Es werden hierbei mehrere Erkennungsalgorithmen kombiniert. Dazu gehören eine Basslinien-Analyse zur Beseitigung von Mehrdeutigkeiten bei der Akkordanalyse und eine Rhythmuserkennung zur Bestimmung von Taktgrenzen, die als Akkordgrenzen angenommen werden. Der Ansatz erreicht eine Trefferquote von durchschnittlich 77,4 %. Der in dieser Arbeit neu vorgestellte Algorithmus HMM/KNN erzielt 80,0 % und damit eine höhere Trefferquote als der Algorithmus von Sumi und Kollegen. Für einen algorithmisch sinnvollen Vergleich sind die Ergebnisse von Bello/Pickens und Lee/Slaney heranzuziehen, da hier Algorithmen der gleichen Komplexität verglichen werden. Hierbei erzielt der neue Algorithmus HMM/KNN eine Steigerung um 5 % auf 80,0 % im Vergleich zum Stand der Technik.

³ International Symposium on Music Information Retrieval 2008.

12.7 Zusammenfassung

In diesem Kapitel wurde ein neuer Algorithmus (HMM/KNN), der HMM mit einem KNN-Wahrscheinlichkeitsschätzer kombiniert, für die Akkorderkennung vorgestellt, evaluiert und mit dem Stand der Technik verglichen. Als Vergleichs- und Evaluierungsgrundlage wurden die Algorithmen KNN und HMM verwendet. Der neue Algorithmus HMM/KNN erreicht für die Beatles-Alben „Please Please Me“ und „Beatles For Sale“ eine Trefferquote von 80,0 % im Mittel. Im Vergleich zum Stand der Technik ist der neue Algorithmus ca. 5 % besser. Darüber hinaus wurden drei verschiedene Merkmalsvektoren evaluiert und verglichen. Der hier neu vorgestellte Merkmalsvektor VSYM, der auf dem Symmetriemodell aufbaut, erreicht in Verbindung mit den Analysealgorithmen bessere Ergebnisse als die Merkmalsvektoren HSG6D und CHROMA. Dies stellte sich gegenüber dem Merkmalsvektor CHROMA als signifikant heraus.

Kapitel 13

Zusammenfassung und Ausblick

13.1 Zusammenfassung

Das Ziel dieser Arbeit bestand darin, Verbesserungen in der Analyse und der Synthese von Audiosignalen durch Anwendung von Tonräumen zu erreichen. Dazu mussten Verfahren entwickelt werden, die es ermöglichen, sowohl Audiosignale in Tonräumen darzustellen und zu analysieren als auch Audiosignale aus Tonräumen heraus zu erzeugen. Um dieses Ziel zu erreichen, wurde zunächst der Stand der Technik von Tonraummodellen analysiert. Ausgehend davon bestand der erste Schritt der vorliegenden Arbeit in der Entwicklung einer mathematischen Beschreibung, welche Tonalität - angelehnt an Lerdahls Idee - auf verschiedenen hierarchischen Ebenen beschreibt. Aufbauend auf David Gatzsches Kadenzkreis wurde die so genannte *Symmetriemodell- Generatorformel* entwickelt, die die Entstehung der vier Hauptebenen der abendländischen Tonalität (die Quintebene, die Akkord- bzw. Terzebene, die diatonische Tonleiterebene und die chromatische Ebene) sehr effektiv beschreibt. Darauf aufbauend kam es zur Entwicklung des *SYM-Operators*, mit dessen Hilfe verschiedene, auf eine gewählte Tonart bezogene Modelle - ebenfalls aus der Quintebene heraus - erzeugt und auf den jeweiligen Symmetrieton zentriert werden. Durch diese Zentrierung gelingt es, unterschiedliche wahrnehmungspsychologische Eigenschaften von Musik numerisch direkt zu repräsentieren. Durch die Entwicklung der beiden Formeln wurde es möglich, eine Vielzahl bereits existierender Modelle zu verbinden und in ein einheitliches, als *Symmetriemodell* bezeichnetes Framework zu integrieren. Ein weiterer Beitrag war die Entwicklung einer schlüssigen Symbolik zur Identifikation der Teilräume des Modells.

Alle der bisher genannten Tonalitätsmodelle erlauben lediglich die Einordnung symbolischer Musiksignale, d.h. Töne haben nur zwei Zustände, sie sind vorhanden oder nicht

vorhanden. Reale Musiksignale dagegen sind durch Obertöne, geräuschhafte Anteile oder sonstige Störungen gekennzeichnet. Um auch derartige Signale mit Hilfe von Tonräumen analysieren zu können, wurde ein Verfahren entwickelt, bei dem die zu einem Zeitpunkt hörbaren Tonigkeiten als komplexe Vektoren im kreisförmigen Tonraum repräsentiert werden. Der Winkel des Vektors ergibt sich dabei aus der Position der Tonigkeit im Tonraum, die Länge des Vektors aus der aktuellen Energie der Tonigkeit. Durch Addition der komplexen Vektoren aller Tonigkeiten kann ein *Summenvektor* berechnet werden, welcher wichtige tonale Eigenschaften eines Musiksignals niedrigdimensional zusammenfasst.

Die musiktheoretische Evaluation des Symmetriemodells und des Summenvektors zeigte, dass das Symmetriemodell beziehungsweise der Summenvektor *musiktheoretische Elemente* korrekt repräsentiert. Neben dem bereits von David Gatzsche herausgearbeiteten Aspekten des Akkordaufbaus und der Funktionstheorie wurden als weitere semantische Eigenschaften die Terminierung des Tongeschlechts und der Haupt- und Nebenfunktionen, die Repräsentation von Charakterdissonanzen, die Repräsentation entfernter Terzverwandtschaften sowie die Repräsentation von Strebefähigkeit, Leit- und Strebetonigkeiten und von Alterationen ermittelt. All diese Eigenschaften werden sehr deutlich durch Teilräume des Symmetriemodells und den darin berechneten Summenvektoren repräsentiert.

Als erste Anwendung des Symmetriemodells wurde ein als *Tonraumsynthesizer* bezeichnetes Verfahren vorgestellt, das die Erzeugung von Musiksignalen auf Basis von Tonräumen ermöglicht. Dadurch, dass der Tonraum sinnvolle Tonkombinationen benachbart anordnet, ist es möglich, mithilfe sehr einfacher geometrischer Auswahlformen (*Selektionsflächen*) wohlklingende Tonkombinationen zu erzeugen. Durch die Transformation des gewählten Tonraumes können Akkorde verändert (zum Beispiel überblendet) werden. Die musikalische Logik des Tonraumes wird somit direkt erfahrbar, hörbar und erlebbar. Da das Symmetriemodell ausschließlich aus chroma- bzw. tonigkeitsorientierten Teilräumen besteht, wurde der so genannte *Tonigkeits-Tonhöhenraum* eingeführt. Dieser ordnet den unterschiedlichen Oktavlagen von Tonigkeiten geometrische Positionen dergestalt zu, dass durch einfaches Verschieben der Selektionsfläche Umkehrungen oder „günstige“ Akkordverbindungen erzeugt werden. Als konkretes Ausführungsbeispiel eines Tonraumsynthesizers wurde das *HarmonyPad* vorgestellt. Das HarmonyPad kombiniert mehrere Teilräume des Symmetriemodells und erlaubt das Spiel unterschiedlichster tonaler Elemente, angefangen bei wohlklingenden Akkordkadenzen bis hin zu komplexen Modulationen. Der Spieler interagiert dabei direkt mit einer Visualisierung des Tonraumes und kann somit sämtliche der zuvor beschriebenen musiktheoretischen Zusammen-

hänge direkt hören, sehen und erfahren. Das Potential des HarmonyPads für neuartige elektronische Spiele wurde darüber hinaus durch Anbindung und Evaluierung verschiedener Spielecontroller untersucht.

Der zweite große Teil der vorliegenden Arbeit bestand in der Anwendung des Symmetriemodells auf konkrete Probleme des Music Information Retrievals, nämlich die Erkennung von Tonarten und Akkorden sowie die Repräsentation von aufgezeichneten Musiksignalen im Symmetriemodell.

Es wurde gezeigt, dass *musikalische Ereignisse* (hier: Audioaufnahmen, reales Audio) – wie z.B. Töne, Intervalle und Akkorde – im Symmetriemodell funktionsrichtig eingeordnet werden. Dabei trägt der Winkel eines Summenvektors in den tonartbezogenen Quint- und Terzkreisen die harmonische Information. Die Länge eines Summenvektors repräsentiert die „Rauschfreiheit“ des Ereignisses und spielt im Vergleich zum Winkel harmonisch gesehen eine untergeordnete Rolle, da z.B. ein verzerrter C-Dur-Gitarrenakkord (rauschhaft) und ein klarer C-Dur-Orgelakkord (rauschfrei) harmonisch identisch sind. Hierbei besitzen die Summenvektoren den gleichen Winkel, haben aber unterschiedliche Längen. Die Auswertung mittels Histogrammen ist geeignet, um ganze Musikstücke zu analysieren und zu bewerten.

Die Grundlagen für das *maschinelle Lernen* wurden erläutert und die beiden Verfahren *Hidden Markov Models* und *K Nearest Neighbours* vorgestellt. Die meisten maschinellen Lernverfahren - wie z.B. *Gaussian Mixture Models*, *Support Vector Machines* und auch *k Nearest Neighbours* - verwenden nur eine funktionelle Relation und klassifizieren jedes Einzelereignis sofort. Bei *Hidden Markov Models* kommt zusätzlich eine zeitliche Relation zum Einsatz, die nicht nur den aktuellen Merkmalsvektor zur Klassifizierung, sondern auch das Klassifizierungsergebnis des zeitlichen Vorgängers benutzt. Für die Schätzung der Wahrscheinlichkeiten werden bei *Hidden Markov Models* generell Gaußmischverteilungen verwendet. In dieser Arbeit wurden die Gaußmischverteilungen durch einen *k Nearest Neighbours*-Schätzer ersetzt. Bei der *Akkorderkennung* wurde für 100 Beatles-Lieder dadurch die Trefferquote an richtig erkannten Akkorden von 69,8 % auf 80,0 % gesteigert. Der Stand der Technik liefert mit einem wesentlich komplexeren Ansatz eine Trefferquote von 77,4 %. Mit dem neuen Ansatz vergleichbare Algorithmen erreichen hierbei nur 75,1 %.

Der Vergleich verschiedener *Merkmalsvektoren* zeigte, dass der Merkmalsvektor des Symmetriemodells in Verbindung mit den angewendeten Maschinenlernverfahren am besten abschneidet und signifikant bessere Ergebnisse erzielt als der Chromavektor, der den

Stand der Technik repräsentiert. Der Merkmalsvektor des Symmetriemodells ist ebenfalls besser geeignet als der 6D-Vektor von Harte, Sandler und Gasser.

Die *Tonarerkennung* verfolgte bisher nur das Ziel, ein Musikstück einer Dur- oder Molltonart zuzuordnen. Der Stand der Technik erreicht bei den verwendeten 442 Teststücken aus den Musikrichtungen *Klassik*, *Beatles* und *RockPop* eine maximale Tonarterkennungsrate von 64,2 %. Die Ergebnisse werden mit dem MIREX-Score bewertet. Der hier neu vorgestellte Ansatz, der mit den Merkmalsvektoren des Symmetriemodells arbeitet, kommt auf 82,9 %. Ein Derivat des Algorithmus von Alexander Lerch erzielt 77,8 % und ein einfacher Algorithmus, der Dur/Moll-Korrelationen berechnet, erreicht 77,6 %. Über diese Tonarterkennung hinaus wurde ein vollkommen neuer Ansatz präsentiert, der *Kirchentonarten* erkennt. Da die Kirchentonarten den Charakter eines Musikstückes wesentlich besser beschreiben als die Unterscheidung in Dur und Moll, ist der neue Ansatz für den Vergleich von Musikstücken besser geeignet als der bisherige. Zur Evaluierung wurde ein neuer Bewertungsmaßstab eingeführt.

13.2 Ausblick

Einer der nächsten Schritte ist, die in dieser Arbeit entwickelten tonraumbasierten Analyse- und Syntheseverfahren zu kombinieren. So wäre es z.B. möglich, Audiosignale als Echtzeitströme oder per Audiodatei mit Hilfe der hier entwickelten Tonart- und Akkorderkennungsalgorithmen zu analysieren und derart im HarmonyPad zu visualisieren, dass es dem Spieler ermöglicht wird, 1.) die in dem Musiksignal enthaltenen Akkorde zu erkennen und 2.) die dazu passenden Akkorde zu spielen.

Ein weiterer wichtiger Forschungspunkt ist die Ableitung zielgruppenspezifischer Ausführungsvarianten des HarmonyPads. Je nachdem ob es sich um Kinder im Vorschulalter, Schüler, Erwachsene, Senioren, Kompositionsstudenten oder DJs handelt, sind ganz unterschiedliche Benutzerschnittstellen zu entwickeln. In Zusammenarbeit mit der Abteilung für Kindermedien des IDMTs werden bereits Schritte in diese Richtung gegangen. So wird z.B. an der Entwicklung einer virtuellen interaktiven Tanzumgebung gearbeitet, welche es mehreren Kindern ermöglicht, durch Ganzkörperinteraktion gemeinsam Musik zu erzeugen. Ein anderer Forschungspunkt ist die Integration weiterer klassischer Stimmführungsregeln, so dass das HarmonyPad z.B. auch das Hören und Komponieren nach den Regeln des strengen vierstimmigen Satzes ermöglicht. Die Weiterentwicklung des HarmonyPads zu einem „akustischen“ Instrument – z.B. durch Kombination mit am Fraunhofer IDMT entwickelten Flachlautsprechertechnologien – ist ein viel ver-

sprechender Forschungsgegenstand. Interessant ist darüber hinaus die Integration des HarmonyPad-Spielprinzips in aktuelle elektronische Musikspiele.

Bei der tonraumbasierten Analyse von Musiksignalen ist in weiteren Forschungen zu untersuchen, inwieweit der neue Ansatz zur Tonarterkennung mittels Kirchentonarten für die Musiksuche oder den Vergleich von Musikstücken geeignet ist, da die Tonart ein Merkmal von vielen ist, das dazu angewendet werden kann. Der Merkmalsvektor des Symmetriemodells ist des Weiteren für die musikalische Ähnlichkeitssuche geeignet. Dies betrifft sowohl die Suche nach Musikstücken mit dem gleichen Musikstil als auch die Suche nach Cover-Versionen oder aber die Plagiatssuche. Bei der Akkorderkennung ist zu evaluieren, ob komplexere Hidden Markov Modelle mit höherer Ordnung bessere Ergebnisse erzielen können. Dabei sind auch voneinander abhängige HMMs vorstellbar, die je nach Zustandsmodellierung (z.B. über die Tonarterkennung oder Instrumentenerkennung) ihre Wahrscheinlichkeitsmodelle anpassen können.

Literaturverzeichnis

- [1] MIR Research, Internet Blog, Tuesday, 25 September 2007, Title: Beatles Chord Transcriptions, <http://mir-research.blogspot.com/2007/09/beatles-chord-transcriptions.html>
- [2] 3DConnexion, A Logitech company: SpaceNavigator. In: <http://www.3dconnexion.com> (2008), Jan
- [3] Webseite der Firma Analog Devices, <http://www.analog.com>. 2008
- [4] Website der WiiLi Entwickler, <http://www.wiili.org/> Wiili. 2008
- [5] The Yamaha TenoryOn. In: Webpage, <http://tenori-on.yamaha-europe.com/uk/> (2009), February
- [6] Vorlesung Musik- und Audioinformatik, WS'06/07. In: http://www.encycloSPACE.org/MusikInformatik_1 (2011)
- [7] ALEXANDER, Peter L.: *How MIDI works*: ISBN: 978-0-634-02083-4. 6. Hal Leonard Publishers Company, 2001
- [8] ALPAYDIN, E.: *Maschinelles Lernen*. Oldenbourg Verlag, 2008. – ISBN 978-3-486-58114-0
- [9] BALZANO, G. J.: The Group-theoretic Description of 12-fold and Microtonal Pitch Systems. In: *Computer Music Journal* 4 (1980), Nr. 4
- [10] BALZANO, G. J.: The Pitch Set as a Level of Description for Studying Musical Pitch Perception. In: *Music, Mind, and Brain*. (1982)
- [11] BELLO, Juan P. ; DAUDET, Lournet ; SAMER, Abdallah ; DUXBURY, Chris ; DAVIES, Mike ; SANDLER, Mark B.: A tutorial on onset detection in music signals: IEEE Transactions on Speech and Audio Processing, 13(5, part 2). (2005), S. 1035–1047

- [12] BELLO, Juan P. ; PICKENS, Jeremy: A Robust Midlevel Representation for Harmonic Content in Music Signals: ISMIR 2005, London, UK. In: *International Symposium on Music Information Retrieval* (2005)
- [13] BOYER, Stephen W.: *Device for patterned input and display of musical notes*. Patent, US 2001/0052283 A1, Dec 2001
- [14] BREGMAN, Albert S.: *Auditory scene analysis*. Second MIT Press paperback edition. Cambridge Mass. : MIT Press, 1999. – ISBN 0–2620–2297–4
- [15] BROWN, Judith C. ; PUCKETTE, Miller S.: An Efficient Algorithm for the Calculation of a Constant Q Transform. In: *Journal of the Acoustic Society of America* (1992)
- [16] BRUNER, Cheryl L.: The Perception of Contemporary Pitch Structures. In: *Music Perception* 2 (1984), Nr. 1, S. 25–39
- [17] BUCHANAN, William J.: *The Handbook of Data Communications and Networks*: ISBN: 978-1-402-07741-8. Bd. 1. 2. Kluwer Academic Publishers, 2004
- [18] BUDGE, H.: A study of chord frequencies. In: *New York: Bureau of Publications, Teachers College, Columbia University* (1943)
- [19] CALLENDER, Clifton ; QUINN, Ian ; TYMOCZKO, Dmitri: Generalized Voice-Leading Spaces. In: *Science* 320 (2008), S. 346–348
- [20] CHEW, Elaine: *Towards a Mathematical Model of Tonality*, Massachusetts Institute of Technology, Diss., 2000
- [21] CHEW, Elaine: Modeling Tonality: Applications to Music Cognition: Wheat Ridge, Colorado. In: *Cognitive Science Society* 23 (2001), S. 206–211
- [22] CHEW, Elaine: *Spiral Array: An Algorithm for Determining Key Boundaries*: Springer Verlag. (2002), S. 18–31
- [23] CHU, Ellen W. ; CHU, Eleanor Chin-Hwa ; GEORGE, Alan: *Inside the FFT Black Box: Serial and Parallel Fast Fourier Transform Algorithms*: ISBN: 978 0 849 30270-1. Crc Pr Inc, 2000
- [24] CHUAN, Ching-Hua ; CHEW, Elaine: Polyphonic Audio Key Finding Using the Spiral Array CEG Algorithm: IEEE 2002, Amsterdam, The Netherlands. In: *International Conference on Multimedia and Expo* (2002)

- [25] COHN, Richard: Neo-riemannian operations, parsimonious trichords, and their tonnetz representations. In: *Journal of Music Theory* 41 (1997), S. 1–66
- [26] COHN, Richard: Introduction to Neo-Riemannian Theory: A Survey and a Historical Perspective. In: *The Journal of Music Theory* 42 (1998), Nr. 2
- [27] COOK, Norman D. ; FUJISAWA, Takashi X.: The Psychophysics of Harmony Perception: Harmony is a Three-Tone Phenomenon. In: *Empirical Musicology Review* 1 (2006), Nr. 2
- [28] CROWDER, Robert G.: Perception of the Major/Minor Distinction: I. Historical and Theoretical Foundations. In: *Psychomusicology* 4 (1985), Nr. 1, S. 3–12
- [29] CROWDER, Robert G.: Perception of the Major/Minor Distinction: II. Experimental Investigations. In: *Psychomusicology* 5 (1985), Nr. 1, S. 3–24
- [30] CROWDER, Robert G.: Perception of the Major/Minor Distinction: III. Hedonic, musical, and affective discriminations. In: *Bulletin of the Psychonomic Society* 23 (1985), Nr. 4, S. 314–316
- [31] CROWDER, Robert G. ; REZNICK, J. S. ; ROSENKRANTZ, Stacey L.: Perception of the Major/Minor Distinction: V. Preferences among infants. In: *Bulletin of the Psychonomic Society* 29 (1991), Nr. 3, S. 187–188
- [32] DEBBELER, Judith: *AESTHETICA THEATRALIA*. Bd. 4: *Harmonie und Perspektive: Die Entstehung des neuzeitlichen abendländischen Kunstmusiksystems*: ISBN: 978-3-940388-02-5. München : epodium Verlag, 2007
- [33] DEUTSCH, Diana ; FEROE, J.: The Internal Representation of Pitch Sequences in Tonal Music. In: *Psychological Review* 88 (1981), S. 503–522
- [34] DORRELL, Philip: *WHAT IS MUSIC?* Revision Date: 22 March 2005. Philip Dorrell, 2005. – ISBN 1–4116–2117–4
- [35] İZMIRLI, Özgür: Audio Key Finding Using Low-Dimensional Spaces: ISMIR, Victoria, Canada. In: *International Conference on Music Information Retrieval* (2006)
- [36] İZMIRLI, Özgür: *Audio Key Finding Using Low-Dimensional Spaces: Proceedings of the International Conference on Music Information Retrieval ISMIR, Victoria, Canada*. 2006
- [37] İZMIRLI, Özgür: *Localized Key Finding from Audio Using Non-negative Matrix Factorization: Proceedings of the International Conference on Music Information Retrieval ISMIR, Vienna, Austria*. 2007

- [38] DROBISCH, Wilhelm M.: *Abhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften*. Bd. 4,[1: *Über musikalische Tonbestimmung und Temperatur*. Leipzig : Hirzel, 1855
- [39] DUXBURY, Chris ; SANDLER, Mark B. ; DAVIES, Mike: *A Hybrid Approach to Musical Note Onset Detection*: in *Proceedings of the 5th International Conference on Digital*. (2002)
- [40] ESSBACH, Anglea: *Theorie und Praxis des Usability Testings – Entwicklung eines Konzepts*, Technische Universität Ilmenau, Diplomarbeit, 2006
- [41] EULER, Leonhard: *Tentamen novae musicae (1739)*. Broude, 1968
- [42] FASTL, Hugo ; ZWICKER, Eberhardt: *Psychoacoustics: Facts and Models*. 3rd edition (December 28, 2006). Springer, 2006. – ISBN 3540231595
- [43] FASTL, Hugo ; ZWICKLER, Eberhard: *Springer Series in Information Sciences*. Bd. 22: *Psycho-Acoustics: Berlin Heidelberg New York, 3. Edition, 2006, ISBN: 978-3-540-23159-2*. 2006
- [44] FLEISCHER, W. ; BARZ, I. ; SCHRÖDER, M.: *Wortbildung der deutschen Gegenwartssprache*. Max Niemayer Verlag, 1992
- [45] FORTE, Alan: *The Structure of Atonal Music*. Yale University Press, 1977. – ISBN 0300021208
- [46] FRESE, Peter: *Akkordharmonisches Arbeitshilfsgerät*. Schutzrecht, DE 37 44 255 A1, 1987
- [47] GATZSCHE, David: *Visualisierung musikalischer Parameter in der Musiktheorie*, Hochschule für Musik Franz Liszt, Diplomarbeit, 2004
- [48] GEORGE, Michael P.: *Method and System of Studying Music Theory*. Schutzrecht, US 2002/0178896 A1, 2002
- [49] GOLLIN, E.: *Some aspects of three-dimensional Tonnetze*. In: *Journal of Music Theory* (1998), S. 195–206
- [50] GOLLIN, Edward H.: *Representations of Space and Conceptions of Distance in Transformational Music Theories*, Harvard University, Diss., 2000
- [51] GOMEZ, E.: *Evaluation and applications of tonal profiles for automatic music tonality description*. In: *9th International Conference on Music Perception and Cognition* (2006)

- [52] GÓMEZ, Emilia: Key Estimation from Polyphonic Audio. In: *MIREX Contest 2005* (2005)
- [53] GREEN, Judith ; D'OLIVEIRA, Manuela: *Learning to use statistical tests in psychology*. Third Edition. Open University PRes, 2006. – ISBN 9780335216802
- [54] HAINSWORTH, Stephen W. ; MACLEOD, Malcolm: Onset Detection in Musical Audio Signals: Singapore. In: *Proceedings of ICMC* (2003)
- [55] HAMANAKA, Masatoshi ; HIRATA, Keiji ; TOJO, Satoshi: ATTA: Implementing GTTM on a Computer. In: *Proceedings of the International Conference on Music Information Retrieval ISMIR* (2007)
- [56] HARTE, Christopher ; SANDLER, Mark: Automatic chord identification using a quantized Chromagram: Audio Engineering Society, Spain. In: *Audio Engineering Society* (2005)
- [57] HARTE, Christopher ; SANDLER, Mark ; ABDALLAH, Samer ; GÓMEZ, Emilia: Symbolic Representation of Musical Chords: A proposed Syntax for Text Annotations. In: *Proceedings of the 6th International Conference on Music Information Retrieval, ISMIR2005, 2005*
- [58] HARTE, Christopher ; SANDLER, Mark ; GASSER, Martin: Detecting Harmonic Change In Musical Audio. In: *AMCMM'06, October 27* (2006)
- [59] HARTE, Christopher ; SANDLER, Mark ; GASSER, Martin: Detecting Harmonic Change in Musical Audio: Santa Barbara, CA. In: *Audio and Music Computing for Multimedia Workshop* (2006)
- [60] HARTE, Christopher ; SANDLER, Mark B.: Automatic Chord Identification using a Quantised Chromagram. In: *Proceedings of the Audio Engineering Society* (2005)
- [61] HARTUNG, Joachim ; ELPELT, Bärbel ; KLÖSENER, Karl-Heinz: *Statistik. Lehr- und Handbuch der angewandten Statistik: ISBN: 3-486-20218-9*. 5. München : Oldenbourg Verlag, 1986
- [62] HAUNSCHILD, Frank: *The New Harmony Book*. Enlarged and revised new 1994 edition. AMA Verlag GmbH, 1994. – ISBN 3-927190-68-3
- [63] HAUPTMANN, Moritz: *Die Natur der Harmonik und Metrik zur Theorie der Musik*. Breitkopf und Härtl, 1853

- [64] HEADLAM, Dave ; BROWN, Matthew: Comment on „The Geometry of Musical Chords“. In: *Science* 315 (2007)
- [65] HEINICHEN, Johann D.: *Der Generalbass in der Komposition*. 2nd Nachdruck der Ausg. Dresden 1728 edition (1994). G. Olms Verlag, 1994. – ISBN 3487098245
- [66] HEINZ, Thorsten: *Ein physiologisch gehörgerechtes Verfahren zur automatisierten Melodietranskription*, Fakultät für Elektrotechnik und Informationstechnik, Technische Universität Ilmenau, Diss., 2004
- [67] HELMHOLTZ, Hermann von: *Die Lehre von den Tonempfindungen*. Neuauflage, Original erschien 1863. Hildesheim : Olms, 2000. – ISBN 3487019744
- [68] HERRMANN, Klaus: *Gerät zur Veranschaulichung der Intervalle, der Tonarten in Dur und Moll, ihrer Tonleitern, Vorzeichnungen und wechselseitigen Beziehungen, zur Ermittlung der Akkorde und zum Transponieren von Tonarten in Dur und Moll*. Schutzrecht, DE 89 02 959 U1, 1989
- [69] HONINGH, Aline: *The Origin and Well-Formedness of Tonal Pitch Structures*, Institut for Logic, Language and Computation, Diss., 2006
- [70] HUTCHINSON, W. ; KNOPOFF, L.: The acoustic component of Western consonance. In: *Interface* 7 (1978), S. 1–29
- [71] HYER, Brian: Re-Imagining Riemann. In: *Journal of Music Theory* 39 (1995), Nr. 1, S. 101–138
- [72] IZMIRLI, Özgür: An Algorithm for Audio Key Finding: MIREX05, London, UK. In: *Music Information Retrieval Evaluation eXchange* (2005)
- [73] JAZZMUTANT. *Website der Firma Jazzmutant*, <http://www.jazzmutant.com/>. 2008
- [74] JETSCHKE, Sebastian ; KOELSCH, Stefan ; FRIEDRICH, Angela D.: Investigating the Relationship of Music and Language in Children, Influences of Musical Training and Language Impairment. In: *Ann. N.Y. Acad. Sci.* 1060 (2005), S. 231–242
- [75] JUNG, Volker: *Handbuch für die Telekommunikation*: ISBN: 978-3-540-42795-7. 2. Berlin : Springer Verlag, 2002
- [76] KAMEOKA, A. ; KURIYAGAWA, M.: Consonance theory Part II: Consonance of complex tones and its calculation method. In: *Journal of the Acoustical Society of America* (1969), Nr. 45, S. 1460–1469

- [77] KASTNER, Marianna P. ; CROWDER, Robert G.: Perception of the Major/Minor Distinction: IV. Emotional Connotations in Young Children. In: *Music Perception* 8 (1990), Nr. 2, S. 198–202
- [78] KELLNER, David: *Treulicher Unterricht im General-Bass*. Christian Herold, 1737
- [79] KOELSCH, Stefan ; SIEBEL, Walter A.: Towards a neural basis of music perception. In: *Trends in Cognitive Sciences* 9 (2005), Nr. 12
- [80] KRUMHANSL, C. ; KESSLER, E.: Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys. In: *Psychological Review* (1982), Nr. 89, S. 334–368
- [81] KRUMHANSL, Carol L.: The psychological representation of musical pitch in a tonal context. In: *Cognitive Psychology* 11 (1979), S. 346–374
- [82] KRUMHANSL, Carol L.: *Cognitive foundations of musical pitch*. Oxford psychology series; no.17. Oxford University Press, 1990. – ISBN 0–19–505475–X
- [83] KRÄMER, Thomas: *Lehrbuch der harmonischen Analyse*. Breitkopf und Härtel, 1997. – ISBN 978–3765103056
- [84] KRÄMER, Thomas: *Harmonielehre im Selbststudium*. Neuausg. Wiesbaden : Breitkopf und Härtel, 2006. – ISBN 9783765102615
- [85] KRÄMER, Thomas ; DINGS, Manfred: *Lexikon Musiktheorie*. Wiesbaden : Breitkopf und Härtel, 2005
- [86] KRÖMKER, Heidi. *Vorlesung Softwareergonomie: Methoden der Evaluation, TU-Ilmenau, 2008*
- [87] KUUSI, Tuire: *Set-Class and Chord: Examining Connection between Theoretical Resemblance and Perceived Closeness*, Sibelius Academy Helsinki, Diss., 2001
- [88] KUUSI, Tuire: Theoretical Resemblance versus Perceived Closeness: Explaining Estimations of Pentachords by Abstract Properties of Pentad. In: *Journal of New Music Research, Volume 31, Issue 4 December 2002* , pages 377 - 388 31 (2002), Nr. 4, S. 377–388
- [89] KÖLSCH, Stefan: Neural substrates of processing syntax and semantics in music. In: *Current Opinion in Neurobiology* (2005), Nr. 15, S. 1–6

- [90] KÖLSCH, Stefan ; MULDER, Juul: Electric brain responses to inappropriate harmonies during listening to expressive music. In: *Clinical Neurophysiology* (2002), Nr. 113, S. 862–869
- [91] KÜFFNER, Peter: *Musik-Hilfsmittel zum Einstellen von Tonart, Tonleiter und entsprechenden harmonischen Zusammenhängen*. Schutzrecht, DE 4002361 A1, 1990
- [92] LEE, Daniel D. ; SEUNG, Sebastian: Learning the Parts of Objects by Non-negative Matrix Factorization. In: *Nature* (1999), Nr. 401, S. 788–791
- [93] LEE, Kyogu: Automatic Chord Recognition from Audio Using Enhanced Pitch Class Profiles. In: *Proceedings of the International Computer Music Conference ICMC* (2006)
- [94] LEE, Kyogu: A Unified System for Chord Transcription and Key Extraction using Hidden Markov Models: ISMIR, Vienna, Austria, 2007. In: *International Conference on Music Information Retrieval* (2007)
- [95] LEE, Kyogu: A System for Automatic Chord Recognition from Audio Using Genre-specific Hidden Markov Models: Proceedings of the International Workshop Adaptive Multimedia Retrieval, Paris, France. In: *International Workshop Adaptive Multimedia Retrieval* (2008)
- [96] LEFÉVRE, Fabrice ; MONTACIÉ, Claude ; CARATY, Marie-José: K-NN Estimator in a HMM-Based Recognition System: NATO ASI, Jersey. In: *Computational Models of Speech Pattern Processing* (1997), S. 169, pp. 96–101
- [97] LEFÉVRE, Fabrice ; MONTACIÉ, Claude ; CARATY, Marie-José: KNN versus Gaussian in a HMM-Based System. In: *European Conference on Speech Communication and Technology* (1997)
- [98] LEFÉVRE, Fabrice ; MONTACIÉ, Claude ; CARATY, Marie-José: A MLE Algorithm for the k-NN HMM System. In: *European Conference on Speech Communication and Technology* (1999), S. 2733–2736
- [99] LEGRANGE, Ulyesse J.: *Music education aid*. Schutzrecht, US 5709552 A, 1998
- [100] LERCH, Alexander: Ein Ansatz zur automatischen Erkennung der Tonart in Musikdateien. In: *Konferenzband der Tonmeistertagung* (2004)
- [101] LERDAHL, Fred: *Tonal pitch space*. Hardcover. Oxford : Oxford University Press, 2001. – ISBN 0–1950–5834–8

- [102] LEVIN, David: *Generalized musical Intervals and transformations*. New Haven : Yale Univ. Pr., 1987
- [103] LINDEMANN, Uwe: *Arbeitsgerät zum Auffinden von Akkorden, Harmonien und Tonarten*. Schutzrecht, DE 8005260 U1, 1980
- [104] MADSEN, Søren T. ; WIDMER, Gerhard: Key-Finding with Interval Profiles: ICMC2007. In: *International Computer Music Conference (2007)*
- [105] MAHER, T.F.: A rigorous test of the proposition that musical intervals have different psychological effects. In: *American Journal of Psychology* (1980), Nr. 93, S. 309–327
- [106] MALMBERG, C. F.: The perception of consonance and dissonance. In: *Psychological Monographs* 25 (1918), S. 93–133
- [107] MARTINS, José António O.: *Dasian, Guidonian, and other Affinity Spaces in Twentieth-Century Music*, University of Chicago, Diss., 2006
- [108] MATTHESON, J.: *Kleine General-Bass-Schule*. Originally published in 1735. Laaber-Verlag, 1980
- [109] MAZZOLA, Guerino: *Die Geometrie der Töne*. Birkhäuser Verlag, 1990. – ISBN 3–7643–2353–1
- [110] MAZZOLA, Guerino: *presto Software Manual 1989-1994*. SToA music, 1994
- [111] MAZZOLA, Guerino ; GÖLLER, Stefan: *The topos of music*. Basel : Birkhäuser, 2002
- [112] MICHELS, Ulrich (Hrsg.): *dtv-Atlas Musik*. 21., durchgesehene und korrigierte Auflage. Deutscher Taschenbuchverlag, 2005. – ISBN 3–423–03022–4
- [113] MILMEISTER, Gérard: *The Rubato Composer Music Software*. Springer, 2009. – ISBN 978–3–642–00147–5
- [114] MILMEISTER, Gérard: *Rubato Composer*. <http://www.rubato.org/>, 2006
- [115] MITCHELL, Thomas: *Machine Learning: Mcgraw-Hill International Edit: ISBN 0-07-115467-1*. McGraw-Hill, 1997
- [116] NINTENDO. *Bild der Achsen der Wiimote*, <http://www.osculator.net/wiki/uploads/Main/pry-wiimote.gif>. 2008
- [117] NINTENDO. *Website der Nintendo Wii*, <http://de.wii.com>. 2009

- [118] NISHIBORI, Yu ; IWAI, Toshio: TENORI-ON. In: *NIME - New Interfaces for Musical Expression, International Conference* (2006)
- [119] NOLL, Thomas. *Ist der Tonraum flach oder gekrümmt?* 2001
- [120] NOLL, Thomas. *A Mathematical Model for Tone Apperception*. Technical University of Berlin. 2003
- [121] PAPADOPOULOS, H el ene ; PEETERS, Geoffroy: Large-Scale Study of Chord Estimation Algorithms Based on Chroma Representation and HMM. In: *Proceedings of the ICCBMI2007* (2007)
- [122] PAUWS, Steffen: Musical Key Extraction from Audio: Barcelona, Spain. In: *Fifth International Conference on Music Information Retrieval* (2004), S. 96–99
- [123] PEETERS, Geoffroy: Musical Key Estimation of Audio Signals based on Hidden Markov Modeling of Chroma Vectors. In: *Proceedings of the DAFx-06* (2006)
- [124] PIERCE, John R.: *The Science of Musical Sound*. Revised Edition. W. H. Freeman and Company, New York, 1992. – ISBN 07167–6005–3
- [125] PISTON, W.: *Harmony*. 4th ed., Originally published, 1941. Norton, 1978
- [126] PLOMB, R. ; LEVELT, W.: Tonal Consonance and Critical Bandwidth. In: *J. Acoust. Soc. Am.* 38 (1965), S. 548
- [127] PURWINS, Hendrik: *Profiles of Pitch Classes; Circularity of Relative Pitch and Key - Experiments, Models, Computational; Music Analysis, and Perspectives*, Technical University of Berlin, Diss., 2005
- [128] P OHLERT, Werner: *Grundlagen Harmonik*. Zimmermann Musikverlag , Schimper, 2000. – ISBN 3–921729–30–0
- [129] RABINER, Lawrence ; JUANG, Biing-Hwang: *Fundamentals of Speech Recognition: Prentice Hall Signal Processing Series: ISBN: 0-13-015157-2*. Prentice Hall, 1993
- [130] RAHN, John: *Basic Atonal Theory*. MacMillan Publishing Company, 1980. – ISBN 0028731603
- [131] RANDEL, Don M. (Hrsg.): *The Harvard Concise Dictionary of Music and Musicians*. The Belknap Press of Havard University, 1999. – ISBN 0–674–00084–6
- [132] RIEMANN, Hugo: Das Problem des harmonischen Dualismus; Ein Beitrag zur  sthetik der Musik. In: *Neue Zeitschrift f ur Musik, No. 1-4* (1905)

- [133] RIEMANN, Hugo: Ideen zu einer 'Lehre von den Tonvorstellungen'. In: *Jahrbuch der Musikbibliothek Peters, Jahrgang 21/22 (1914/15)*. Rudolf Schwartz, 1916, S. 1–26
- [134] RIEMANN, Hugo: *Handbuch der Harmonielehre*. 8. und 9. Auflage. Breitkopf und Härtel, Leipzig, 1921
- [135] ROADS, Curtis: *The Computer Music Tutorial*: MIT Press, ISBN: 978-0-262-68082-0. In: *The MIT Press* (1996)
- [136] ROHWER, Klaus C.: *Hilfsmittel zur Zuordnung der Töne einer Tonleiter zu einer musikalischen Akkord- oder Skalenbezeichnung (Skalenuhr) und Verfahren zur Konstruktion desselben*. Schutzrecht, DE 19831409 A1, 1983
- [137] SAILER, Christian ; ROSENBAUER, K: A Bottom-Up Approach to Chord Detection. In: *Proceedings of the International Computer Music Conference ICMC2006* (2006)
- [138] SAMPLASKI, Arthur G.: *A Comparison of Perceived Chord Similarity and Predictions of Selected Twentieth-Century Chord-Classification Schemes, Using Multidimensional Scaling and Clustering Techniques.*, Indiana University, Diss., 2000
- [139] SANDLER, Katy Noland M.: Key Estimation Using a Hidden Markov Model. In: *Proceedings of the 7th International Conference on Music Information Retrieval, ISMIR2006*, 2006
- [140] SAPP, Craig S.: Harmonic Visualizations of Tonal Music. In: *Proceedings of the International Computer Music Conference* (2001)
- [141] SCHELP, Sarah: Macht Musik! In: *Die Zeit*, 30.11.2006 (2006), Nr. 49
- [142] SCHUBERT, Emery: *Measurement and Time Series Analysis of Emotion in Music*, University of New South Wales, Diss., 1999
- [143] SCHULTZ, Ingmar: *Lernhilfe zur Musiktheorie*. Schutzrecht, DE 20301012 U1, 2003
- [144] SCHÖNBERG, Arnold: *Structural functions of harmony*. 2. rev. ed. with corr, originally published in 1954. London : Benn, 1969
- [145] SCHÖNBERG, Arnold: *Harmonielehre*. Fotomechanischer Nachdruck der 3. Auflage Wien 1922. Edition Peters, 1977
- [146] SEFFEN, Holger: *Anzeigevorrichtung, insbesondere für harmonische Ton-Intervallstrukturen*. Schutzrecht, DE 19859303 A1, 1985

- [147] SEFFEN, Holger: *Anzeigevorrichtung, insbesondere für harmonische Ton-Intervallstrukturen*. Schutzrecht, DE 29801154 U1, 1998
- [148] SEHE, Alexander ; ELLIS, Daniel P.: Chord Segmentation and Recognition Using EM-trained Hidden Markov Models: Baltimore, MD. In: *International Symposium on Music Information Retrieval* (2003)
- [149] SETHARES, William A.: *Tuning, Timbre, Spectrum, Scale*. Second Edition. Springer-Verlag London Limited, 2005. – ISBN 9781846281136
- [150] SEZNEC, Yann. *Website der Wii Loop Machine*. <http://www.vimeo.com/155978/>. 2007
- [151] SHENOY, Arun ; WANG, Ye: Key, Chord, and Rhythm Tracking of Popular Music Recordings: Computer Music Journal 29(3). In: *Computer Music Journal* 2005, S. 75–86
- [152] SHEPARD, Roger: Geometrical approximations to the structure of musical pitch. In: *Psychological Review* 89(4) (1982), Jul, S. 305–333
- [153] SHEPARD, Roger N.: Geometrical approximations to the structure of musical pitch. In: *Psychological Review* (1982), Nr. 89, S. 305–333
- [154] STEYER, Gottfried: *Betrachtungen zur Mass- und Zahlenordnung des musikalischen Tonmaterials*. 1. Aufl. 09.2002. Peter Lang Frankfurt, 2002. – ISBN 978-3-631-39864-7
- [155] STUMPF, Carl: *Tonpsychologie*. Hirzel, 1883
- [156] STÖCKLMEIER, Christian: *Entwicklung einer auf Hardware Controllern basierten Steuerung zur Navigation in Tonräumen*, Technische Universität Ilmenau, Diplomarbeit, 2008
- [157] SUMI, Kouhei ; ITOYAMA, Katsutoshi ; YOSHII, Kazuyoshi ; KOMATANI, Kazunori ; OGATA, Tetsuya ; OKUNO, Hiroshi G.: Automatic Chord Recognition Based on Probabilistic Integration of Chord Transition and Bass Pitch Estimation: ISMIR, Philadelphia, USA. In: *International Conference on Music Information Retrieval* (2008)
- [158] TEMPERLEY, D.: *The Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press, 2001
- [159] TERHARDT, Ernst: *Akustische Kommunikation*. Gebundene Ausgabe 1998. Berlin : Springer, 1998. – ISBN 3540634088

- [160] THALMANN, Florian: *Musical Composition with Grid Diagrams of Transformations*. Philosophisch-naturwissenschaftliche Fakultät der Universität Bern, 2007
- [161] TOIVAINEN, Petri: Visualization of Tonal Content in the Symbolic and Audio Domains. In: *Computing in Musicology* 15 (2007), S. 73–98
- [162] TOYODA, Shinichiro: Sensillum: an improvisational approach to composition. In: *NIME - New Interfaces for Musical Expression, International Conference, 2007*
- [163] TYMOCZKO, Dmitry: Scale Networks and Debussy. In: *Journal of Music Theory* 48 (2004), Nr. 2, S. 219–294
- [164] VIAL, F.-G.: *Arbre généalogique de l'harmonie*. (1767)
- [165] WARREN, J. D.: Separating pitch chroma and pitch height in the human brain. In: *Proceedings of the National Academy of Sciences of the United States of America*, 2003, S. 10038–10042
- [166] WEBER, Gottfried: *Versuch einer geordneten Theorie der Tonsetzkunst zum Selbstunterricht*. Zweite durchaus umgearbeitete Auflage. Mainz : Schott, 1821
- [167] WILLIAMS, J. K.: An Interactive, Multimedia Environment for Exploring Tonal Pitch Space. In: *International Conference Interactive Computer Aided Learning* (2007)
- [168] ZHU, Yongwei ; KANKANHALLI, Mohan S. ; SHENG, Gao: Music Key Detection for Musical Audio: IMMC, Melbourne, Australia. In: *11th International Multimedia Modeling Conference* (2005)

Abbildungsverzeichnis

2.1	Neurokognitives Modell der Musikwahrnehmung [79].	10
2.2	(a) Drobischs geometrisches Modell zur Veranschaulichung der Begriffe Tonigkeit und Oktavlage: Der in der XY-Ebene dargestellte Kreis reprä- sentiert die Zirkularität der Tonigkeit, die Z-Achse die Tonhöhe. (b), (c), (d) Roger Shepard erweiterte Drobischs Modell um zwei weitere Dimen- sionen, welche die Quintverwandtschaft von Tonigkeiten repräsentieren sollen (aus [152]).	14
2.3	“Die chromatische Tonskala der abendländischen Musik und die Arten der Tonbezeichnung. Obere Reihe: Tonindex. Unter dem Manual: Europäisch traditionelle Tonbezeichnung. Unterste Reihe: Bevorzugte internationale Tonbezeichnung“ [159, S. 202]	15
2.4	Zusammenhang zwischen wahrgenommener Intervallkonsonanz und Fre- quenzgruppenbreite	17
2.5	Carol L. Krumhansls Testtonprofile für (a) C-Dur und (b) c-Moll [80]	30
2.6	Wahrgenommene Nähe von Tönen im C-Dur Kontext (nach [81])	31
2.7	Zweidimensionale Darstellung der vierdimensionalen MDS-Lösung zur Ähnlichkeit von Tonarten (nach [82])	32
3.1	Erste Ansätze zur geometrischen Repräsentation von Modulationen	37
3.2	Anordnung von Tonarten im Vial/Weber/Schönberg-Netz [166]	38
3.3	Der Eulertonraum [109, S. 28]	39
3.4	Das Tonnetz, übernommen aus [49] und angepasst an aktuelle deutsche Begriffsgebung der Funktionstheorie [84]. Die Buchstaben <i>P</i> , <i>G</i> und <i>V</i> re- präsentieren die Verwandtschaftsgrade Paralleldreiklang, Gegendreiklang und Variantdreiklang. Die graue Fläche markiert die leitereigenen Drei- klänge des diatonischen Tonsets C-Dur (ohne verminderten Dreiklang) . .	41
3.5	Unter Annahme von Oktav- und enharmonischer Äquivalenz kann das Tonnetz als ein Torus dargestellt werden (angepasst aus [58]).	43

3.6	Elaine Chews Spiral Array für Durtonarten (nach [20])	44
3.7	Die baumförmige Verknüpfung der Teilräume von Fred Lerdahls Tonal Pitch Space	46
3.8	Der Basic Space für den Akkord C-Dur [101, S. 47]	47
3.9	(a) Der Diatonic Circle of Fifths für C-Dur (nach [101, S. 55]) ordnet Tonigkeiten in Quintverwandtschaft an. (b) Der Common Tone Circle für C-Dur (nach [101, S. 56] ordnet die Grundtöne leitereigener Dreiklänge (im Beispiel C-Dur) entsprechend der Anzahl gemeinsamer Tonigkeiten auf einem Kreis an.	48
3.10	Der Chordal Space, eine zweidimensionale Darstellung der Kombinationen aus Abbildung 3.9a und Abbildung 3.9b (nach [101])	48
3.11	(a) Das Harmonische Heptagon [34], (b) Der Kadenzkreis [47]	53
3.12	Darstellung verschiedener Äquivalenzrelationen im Raum \mathbb{Z}^2 . Callenders, Quinns und Tymoczkos Idee besteht darin, den Raum derartig zu transformieren, dass äquivalente Punkte zusammenfallen.	56
3.13	(a) Die Punkte repräsentieren äquivalente Dreitonkombinationen unter Annahme von Transpositionsäquivalenz. (b) Die Punkte repräsentieren äquivalente Akkorde unter Annahme von Transpositions- und Inversionsäquivalenz [19]. Die Zahlen 0, 1, 2, . . . repräsentieren die Tonigkeiten <i>c, cis, d</i>	57
3.14	(a) Der Dasian Space, (b) Der Guidonian Space [107]	60
3.15	Der Music Designers Z71 [6]	63
3.16	Der Rubato Composer [6]	64
3.17	Das TENORI-ON [5]	65
3.18	Beispiel einer Lernschablone [143]	65
3.19	Tastenlayout des Eulernetz basierten Musikinstrumentes (nach [13])	66
4.1	Beispiel eines kreisförmigen Tonraumes	69
4.2	Der tonartübergreifende Quintenzirkel F	73
4.3	Extraktion einer tonartbezogenen Quintreihe aus der Quintebene	74
4.4	Der tonartbezogene Quintenzirkel <i>FR</i> für das diatonische Tonset (a) C-Dur und (b) A-Dur	74
4.5	Entstehung der Terzreihe aus der Quintreihe	75
4.6	Der tonartübergreifende Terzkreis T	76
4.7	Der tonartbezogene Terzkreis <i>TR</i> für die diatonischen Tonsets (a) C-Dur und (b) A-Dur.	76
4.8	Entstehung der diatonischen Tonleiterreihe aus der Terzreihe	77

4.9	Entstehung der chromatischen Reihe aus der diatonischen Tonleiterreihe	77
4.10	Die Terzspirale	78
4.11	Erweiterung des TRs zur Repräsentation nichtdiatonischer Tonigkeiten .	79
4.12	Erzeugung verschiedener praxisrelevanter kreisförmiger Tonräume durch Anwendung des Modulo-Operators auf eine Quintreihe	82
4.13	Das Symmetriemodell im Überblick	85
5.1	Der Summenvektor als wichtiges Grundelement der Analyse musikalischer Elemente im Symmetriemodell	89
5.2	Repräsentation der Tonigkeit d im tonartübergreifenden Terzkreis (schwarze Pfeile) und im tonartübergreifenden Quintenzirkel F (grauer Pfeil . . .	90
5.3	Repräsentation benachbarter Quintintervalle in den Kreisen FR , TR und DR : Je näher die beiden Tonigkeitsvektoren eines Intervalls im Tonraum beieinander sind, desto eindeutiger kann das Intervall von anderen unterschieden werden.	92
5.4	Längen der Summenvektoren von Kombinationen aus drei Tonigkeiten in den Teilräumen des Symmetriemodells	93
5.5	Darstellung des C-Dur-Dreiklages im Symmetriemodell.	96
5.6	Repräsentation des diatonischen Tonsets C-Dur (graues Segment) und F-Dur (gestricheltes Segment) in den Teilräumen F , T und D des Symmetriemodells (v. l. n. r)	97
5.7	Entwicklung der Länge des Summenvektors im F bei hinzukommenden Tonigkeiten. Die gestrichelte vertikale Linie trennt diatonische und nichtdiatonische Tonigkeiten.	98
5.8	Haupt- und Nebendreiklänge werden (a) im tonartbezogenen Quintenzirkel FR getrennt und (b) im tonartbezogenen Terzkreis TR gruppiert (Beispiel für C-Dur). Aus Gründen der Übersichtlichkeit wurden nicht alle Funktionssymbole eingezeichnet.	99
5.9	Tonika-, Subdominant und Dominantbereich im tonartbezogenen Terzkreis: (a) für C-Dur und (b) für a -Moll	100
5.10	Repräsentation leitereigener und entfernter Terzverwandtschaften in der Terzspirale am Beispiel der Parallelverwandtschaften des Dreiklages C-Dur	103
5.11	Im TR wird die Strebefähigkeit von Tonigkeiten geometrisch einfach veranschaulicht: 1.) Strebetonigkeiten liegen im Rücken des TRs, 2.) Strebetonigkeiten streben in die im TR gegenüberliegende Tonigkeit.	104

5.12	Wichtige Alterationen ergänzen die bestehenden Strebetonigkeiten eines diatonischen Tonsets (hell gestrichelte Pfeile) um weitere Strebetonigkeiten (fette, schwarz gestrichelte Pfeile)	105
5.13	Repräsentation des verminderten Dreiklages in den Teilräumen des Symmetriemodells: (a) <i>F</i> , (b) <i>FR</i> , (c) <i>TR</i> und (d) <i>DR</i>	108
5.14	Repräsentation des Dominantseptakkordes im Symmetriemodell: (a) im tonartübergreifenden Quintenzirkel <i>F</i> , (b) im tonartbezogenen Quintenzirkel <i>FR</i> , (c) im tonartbezogenen Terzkreis <i>TR</i> und (d) im tonartbezogenen diatonischen Tonleiterkreis <i>TR</i>	110
5.15	Umdeutungsmöglichkeiten des Klages <i>G-h-d-F</i> : (a) als Dominantseptakkord in <i>c</i> -Moll, (b) als übermäßiger Sextakkord („verminderter Septakkord auf der tiefalterierten Quinte“) in <i>Fis</i> -Dur [83, S. 48], (c) als verminderter Septakkord auf der tiefalterierten Dominantseptime in <i>dis</i> -Moll [83, S. 52] und (d) als Dursubdominante mit hochalterierter Sexte in <i>D</i> -Dur [83, S. 51]	111
5.16	Repräsentation von Zwischendominanten im <i>TR</i> : (a) <i>C</i> -Dur ⁷ (b) <i>D</i> -Dur ⁷	112
5.17	Repräsentation des verminderten Septakkordes im tonartübergreifenden Quintenzirkel <i>F</i> : Dick durchgezogenes Kreuz: <i>h-d-f-as/gis</i> ; gestricheltes Kreuz: <i>e-g-b-des/cis</i> ; gepunktetes Kreuz: <i>a-c-es-ges/fis</i>	114
5.18	Bei der Einordnung des verminderten Septakkordes in den <i>TR</i> muss zwischen Dur- (a) und Molltonarten (b) unterschieden werden.	115
5.19	Darstellung des übermäßigen Dreiklages (a) im <i>F</i> und (b) im <i>TR</i>	116
5.20	Einordnung der Summenvektoren der Tonikadreiklänge der sieben Kirchentonarten in das Symmetriemodell: (a) <i>FR</i> , (b) <i>TR</i>	117
5.21	Vergleich von Carol Krumhansls empirisch ermittelten Tonstabilitäten (Abschnitt 2.7.1) mit den im Symmetriemodell ermittelten Abstands-Profilen	120
5.22	Vergleich von Carol Krumhansls Testtonprofilen (Abschnitt 2.7.1) mit den Vorhersagen des Symmetriemodells: Gemittelte Dur- und Moll-Testtonprofile	121
6.1	Einfaches Modell zur Veranschaulichung musikalischer Tonvorstellung und Interaktion	124
6.2	Das HarmonyPad	126
6.3	Der Tonigkeits-Tonhöhenraum am Beispiel des <i>TRs</i>	128
6.4	Die Parameter der Tonselektionsfunktion	128
6.5	Erzeugen einer Akkordkadenz durch Verändern des Startwinkels	129
6.6	Der Öffnungswinkel definiert die Anzahl der gleichzeitig hörbaren Tonigkeiten	130

6.7	a) Eine Veränderung des Startradius blendet einen Akkord durch verschiedene Umkehrungen. (b) Eine Erweiterung des Öffnungsradius fügt einem Akkord zusätzliche Oktavtöne hinzu.	130
6.8	Oberflächenlayout des HarmonyPads auf dem Lemur: 1 - Tonigkeitslinie, 2 - einer Tonigkeit zugeordnete Tonerhöhungs- und -erniedrigungstasten, 3 - absolute Tonartwechseltasten, 4 - relative Tonartwechseltasten, 5 - Tonerhöhungstasten, 6,7 - Berührungspunkt auf der Touchoberfläche, 8 - Multitouch-Eingabefläche, 9, 10 - Tonauswahlflächen	134
6.9	Die sechs Freiheitsgrade des 3DConnexion SpaceNavigators [2].	137
6.10	Die Wiimote mit eingezeichneten Bewegungsachsen: 1 - Steuerkreuz, 2 - Taste A [116]	138
6.11	Dauer der Durchführung von Task 1 in Sekunden [156]	142
6.12	Dauer der Durchführung von Task 4 in Sekunden [156]	143
6.13	Wie gut wurden die Töne im Tonraum getroffen? [156]	144
6.14	Wie leicht ist das Einstellen des Öffnungswinkels? [156]	144
6.15	Wie leicht ist der Controller zu erlernen? [156]	145
6.16	Mit Hilfe welcher Klänge begleiteten die Testpersonen den Ton C [156] . .	146
7.1	Ablauf Musikanalyse	150
7.2	Einzelton und dessen Spektren zu verschiedenen Zeitpunkten	151
7.3	Frontend – Berechnen der Chromavektoren	153
7.4	Parameter des zu analysierenden Zeitfensters	154
7.5	Constant-Q-Transformation	155
7.6	Tuning mit 36 Tönen pro Oktave	157
7.7	Mapping des Tonvektors auf den Chromavektor	158
8.1	Überwachtes Training mit geordneten Paaren	163
8.2	Klassifizierung mit KNN	166
8.3	Beispiel für eine Markov-Kette	169
8.4	HMM Modellparameter	172
8.5	HMM – typische Anwendungsfälle und Berechnungsmethoden	173
8.6	Berechnen der Zustandswahrscheinlichkeitsmatrix B	174
8.7	Trellis-Diagramm bei der Viterbi-Dekodierung	174
8.8	HMM - Dekodierung	175
9.1	Diatonisches Tonartprofil C-Dur	178
9.2	Krumhansl/Kessler-Profile für (a) C-Dur und (b) c-Moll	179

9.3	Temperley-Profile für (a) C-Dur und (b) c-Moll	180
9.4	Tonarterkennung mit symbolischem Audio	182
9.5	Tonarterkennung mit realem Audio	185
10.1	Die 37 verwendeten Symmetriemodell-Kreise	195
10.2	Repräsentation des Ereignisses a-Moll in den Kreisen von C-Dur	197
10.3	Einzelton c, Orgel-Klang; (a) Signal im zeitlichen Verlauf; (b) Winkelverlauf in FR, TR und DR; (c) Betragsverlauf in FR, TR und DR; (d) Punktwolken in FR, TR und DR;	199
10.4	Intervall Terz c-e, Orgel-Klang; (a) Signal im zeitlichen Verlauf; (b) Winkelverlauf in FR, TR und DR; (c) Betragsverlauf in FR, TR und DR; (d) Punktwolken in FR, TR und DR	200
10.5	Intervall Quinte c-g, Orgel-Klang; (a) Signal im zeitlichen Verlauf; (b) Winkelverlauf in FR, TR und DR; (c) Betragsverlauf in FR, TR und DR; (d) Punktwolken in FR, TR und DR	201
10.6	Akkord a-Moll, Orgel-Klang; (a) Signal im zeitlichen Verlauf; (b) Winkelverlauf in FR, TR und DR; (c) Betragsverlauf in FR, TR und DR; (d) Punktwolken in FR, TR und DR	203
10.7	Kadenz in C-Dur, Orgel-Klang; (a) Signal im zeitlichen Verlauf; (b) Winkelverlauf in FR, TR und DR; (c) Betragsverlauf in FR, TR und DR; (d) Punktwolken in FR, TR und DR	204
10.8	J. S. Bach „C-Dur Präludium“, 90 sek., Histogramme und Punktwolken	207
10.9	L. van Beethoven „Für Elise“, 90 sek., Histogramme und Punktwolken	208
10.10	W. A. Mozart „Eine kleine Nachtmusik“, 90 sek., Histogramme und Punktwolken	209
11.1	Blockschaltbild Tonarterkennung Dur/Moll; (A): Tonsetenergie, (B): Winkelwichtung, (C): Summenvektorbetrag im TR, (D): Tonsetprofil, (E): Dur-/Moll-Bestimmung im FR	212
11.2	Tonsetprofil; „Any time at all“, Beatles	213
11.3	Tonsetenergie und Energieschwelle; „Any time at all“, Beatles	214
11.4	C-Dur als Dominante, Tonika und Subdominante	215
11.5	Winkelwichtung; Winkelgewicht = 0,6	215
11.6	Winkelwichtung; Winkelgewicht = 0,6; „Any time at all“, Beatles	216
11.7	Betragsverlauf in TR; „Any time at all“, Beatles	217
11.8	Dur/Moll-Bestimmung; „Any time at all“, Beatles	218
11.9	Parameteroptimierung Energieschwelle	219

11.10	Parameteroptimierung Winkelgewicht	219
11.11	Ergebnisse Tonarterkennung; Alle Stücke; Friedman-Test; $\alpha = 5\%$	223
11.12	Ergebnisse Tonarterkennung; Beatles; Friedman-Test; $\alpha = 5\%$	224
11.13	Ergebnisse Tonarterkennung; RockPop; Friedman-Test; $\alpha = 5\%$	224
11.14	Ergebnisse Tonarterkennung; Klassik; Friedman-Test; $\alpha = 5\%$	225
11.15	Blockschaltbild Erkennung Kirchentonarten	226
11.16	Kirchentonarten im TR und FR; Beispiel TR(C), FR(C)	227
12.1	Übersicht Tonraummodelle und Analysealgorithmen	232
12.2	Anpassung der Ground Truth an die Testumgebung; „A hard day’s night“ von den Beatles	233
12.3	Berechnung der Trefferquote; „A hard day’s night“ von den Beatles	234
12.4	Symmetriemodell; (a) nicht normierte Vektoren; (b) normierte Vektoren	237
12.5	Ergebnisse KNN ; Friedman-Test; $\alpha = 5\%$	239
12.6	Ergebnisse HMM ; Friedman-Test; $\alpha = 5\%$	241
12.7	Ergebnisse HMM/KNN ; Friedman-Test; $\alpha = 5\%$	245
12.8	Ergebnisse Akkorderkennung; Überblick ; Friedman-Test; $\alpha = 5\%$	245
12.9	Ergebnisse Akkorderkennung; Vektoren ; Friedman-Test; $\alpha = 5\%$	246
12.10	Ergebnisse Akkorderkennung; Algorithmen ; Friedman-Test; $\alpha = 0,3\%$	247
A.1	Die Längen der Summenvektoren von Intervallen in den Teilräumen des Symmetriemodells	3
C.1	Einordnung der Funktionssymbole in die Terzspirale	12

Tabellenverzeichnis

2.1	Identifikation von Tönen in Form von Tonindizes n_c und Tonigkeitsnamen	15
2.2	Zuordnung von Halbtonabständen HTA, Frequenzverhältnissen FV und Intervallbezeichnungen IVB	16
2.3	Zuordnung von Tonsetindizes n_k zu diatonischen Tonsets K	25
2.4	Die sieben Modi aufbauend auf dem diatonischen Tonset C-Dur (nach [62, S. 71])	27
4.1	Zuordnung von erweiterten Indizes n_C zu Tonigkeiten.	71
4.2	Zusammenfassung aller Begriffe und mathematischen Symbole, um kreisförmige Tonräume zu beschreiben	71
6.1	Navigation im Tonraum durch Verwendung des 3DConnexion SpaceNavigators	136
6.2	Verschiedene Tonart- bzw. Spiralwindungswechsel	137
9.1	Übersicht Tonartprofile für die Tonigkeit c	181
11.1	MIREX-Score	218
11.2	Ergebnisse der Dur/Moll-Evaluierung nach den ersten 30 Sekunden; Parameterevaluierung und Tonarterkennungsraten	220
11.3	Ergebnisse Tonarterkennung nach MIREX-Score; Mittelwerte	222
11.4	Bewertungsmaßstab für die Erkennung von Kirchentonarten	228
11.5	Ergebnisse der Kirchentonarten-Evaluierung; nach den ersten 30 Sekunden; Parameterevaluierung und Tonarterkennungsraten	229
11.6	Ergebnisse der Kirchentonarten-Evaluierung; nach den ersten 30 Sekunden; Erkennungsraten getrennt für Tonset und Tonika; Energieschwelle 0,89	230

12.1	KNN+CHROMA; Evaluierung Abstandsmaß; Trefferquote in Prozent; $TP = 100$; $K/TP = 15$; Training=B15...B100; Test=B1...B14 (Beatles-Album „Please Please Me“)	235
12.2	KNN+CHROMA; Evaluierung TP und K/TP ; Manhattan-Distanz; Trefferquote in Prozent; Training=B15...B100; Test=B1...B14 (Beatles-Album „Please Please Me“)	236
12.3	KNN+HSG6D; Evaluierung Abstandsmaß; Trefferquote in Prozent; $TP = 100$; $K/TP = 15$; Training=B15...B100; Test=B1...B14 (Beatles-Album „Please Please Me“)	236
12.4	KNN+HSG6D; Evaluierung TP und K/TP ; Manhattan-Distanz; Trefferquote in Prozent; Training=B15...B100; Test=B1...B14 (Beatles-Album „Please Please Me“)	236
12.5	KNN+VSYM; Evaluierung Abstandsmaß und Symmetriemodell-Kreise; Vektoren <u>normiert</u> ; Trefferquote in Prozent; $TP = 100$; $K/TP = 15$; Training=B15...B100; Test=B1...B14 (Beatles-Album „Please Please Me“)	238
12.6	KNN+VSYM(FR+TR+DR); Vektoren <u>nicht normiert</u> ; Evaluierung TP und K/TP ; Euklidische Distanz; Trefferquote in Prozent; Training=B15...B100; Test=B1...B14 (Beatles-Album „Please Please Me“)	238
12.7	KNN+VSYM(FR+TR+DR); Vektoren <u>normiert</u> ; Evaluierung TP und K/TP ; Treffer in Prozent; Training=B15...B100; Test=B1...B14 (Beatles-Album „Please Please Me“)	239
12.8	Ergebnisse KNN ; Trefferquote in Prozent; $TP = 500$; $K/TP = 3$;	239
12.9	HMM+VSYM; Evaluierung Symmetriemodell-Kreise; Trefferquote in Prozent; Training=B15...B100; Test=B1...B14 (Beatles-Album „Please Please Me“)	240
12.10	Ergebnisse HMM ; Trefferquote in Prozent	241
12.11	HMM/KNN+CHROMA; Evaluierung Abstandsmaß; Trefferquote in Prozent; $TP = 100$; $K/TP = 15$; Training=B15...B100; Test=B1...B14 (Beatles-Album „Please Please Me“)	242
12.12	HMM/KNN+CHROMA; Evaluierung TP und K/TP ; Kosinus-Distanz; Trefferquote in Prozent; Training=B15...B100; Test=B1...B14 (Beatles-Album „Please Please Me“)	242
12.13	HMM/KNN+HSG6D; Evaluierung Abstandsmaß; Trefferquote in Prozent; $TP = 100$; $K/TP = 15$; Training=B15...B100; Test=B1...B14 (Beatles-Album „Please Please Me“)	243

12.14	HMM/KNN+HSG6D; Evaluierung TP und K/TP ; Euklidische Distanz; Trefferquote in Prozent; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)	243
12.15	HMM/KNN+VSYM; Evaluierung Abstandsmaß und Symmetriemodell-Kreise; Trefferquote in Prozent; $TP = 100$; $K/TP = 15$; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)	243
12.16	HMM/KNN+VSYM(FR+TR); Vektoren <u>normiert</u> ; Evaluierung TP und K/TP ; Kosinus-Distanz; Trefferquote in Prozent; Training=B15... B100; Test=B1... B14 (Beatles-Album „Please Please Me“)	244
12.17	Ergebnisse HMM/KNN ; Trefferquote in Prozent	244
12.18	Ergebnisse Akkorderkennung im Durchschnitt; Merkmalsvektoren ; Trefferquote in Prozent	246
12.19	Ergebnisse Akkorderkennung im Durchschnitt; Analysealgorithmen ; Trefferquote in Prozent	247
12.20	Ergebnisse Akkorderkennung; Vergleich mit Stand der Technik: Harte/Sandler [56], Bello/Pickens [12], Lee/Slaney [94] und Sumi und Kollegen [157]; Trefferquote in Prozent	248
A.1	Vergleich der im Symmetriemodell berechneten Stabilität von Tönen in der Tonart C-Dur mit den von Carol L. Krumhansl gefundenen C-Dur-Profilen (Abbildung 2.5)	4
A.2	Vergleich der im Symmetriemodell berechneten Stabilität von Tönen in der Tonart a-Moll mit den von Carol L. Krumhansl gefundenen a-Moll-Profilen (Abbildung 2.5)	5
D.1	Ergebnismatrix - Zusammenfassen mehrerer Ergebnislisten	14
D.2	Rangmatrix	14
E.1	<i>Beatles</i> -Stücke und deren Tonarten – Teil 1	18
E.2	<i>Beatles</i> -Stücke und deren Tonarten – Teil 2	19
E.3	<i>RockPop</i> -Stücke und deren Tonarten – Teil 1	20
E.4	<i>RockPop</i> -Stücke und deren Tonarten – Teil 2	21
F.1	Ergebnisse Akkorderkennung – Teil 1	24
F.2	Ergebnisse Akkorderkennung – Teil 2	25
F.3	Ergebnisse Akkorderkennung – Teil 3	26

Glossar und Abkürzungen

12-D	12-dimensional
A	Zustandsübergangsmatrix
a_{ij}	Zustandsübergangswahrscheinlichkeit von Zustand i nach Zustand j
α	Anzahl der Unterteilungen pro Oktave
α	Winkel der Realtonigkeit im kreisförmigen Tonraum m
$argmax$	Maximierung des angegebenen Parameters (Argument)
B	Zustandswahrscheinlichkeitsmatrix
b	Wahrscheinlichkeitsschätzer
β	Stauchungsfaktor
Bit	Binary Digit
C	Tonartübergreifender chromatischer Kreis (Key Spanning Chromatic Circle)
C_i	Klasse i
CEG	Center of Effect Generator
CD	Compact Disc
CHROMA	Chromavektor
CR	Tonartbezogener chromatischer Kreis (Key R elated C hromatic Circle)
CV	Chromavektor
D	Tonartübergr. diaton. Tonleiterkreis (Key Spanning Circle of D iatonic Scales)
D	Dominante
$d(x, y)$	Metrik
DR	Tonartbezog. diaton. Tonleiterkreis (Key R elated Circle of D iatonic Scales)
e	Eulersche Zahl
$E(\cdot)$	Gesamtverlust
EEG	Elektroenzephalogramm
EKP	Ereigniskorreliertes Potential
EM	Expectation Maximization; Erwartungswertmaximierung
EPCP	Enhanced Pitch Class Profile
ERAN	Early Right Anterior Negativity

<i>F</i>	Tonartübergreifender Quintenzirkel (Key Spanning Circle of Fifth)
f_{cq}	Mittelfrequenz der Constant-Q-Filter
f_{min}	Kleinste Frequenz
FFT	Fast Fourier Transformation
FR	Tonartbezogener Quintenzirkel (Key R elated Circle of F ifth)
FV	Frequenzverhältnis
G	Gegenklang
g	Halbtonauflösung des kreisförmigen Tonraums
$g(\cdot)$	Modell der Hypothesenklasse
GMM	Gaussian Mixture Models
H	Hierarchie
HMM	Hidden Markov Models
HMM/KNN	Hidden Markov Models mit KNN-Wahrscheinlichkeitsschätzer
HTA	Halbtonabstand
HPCP	Harmonic Pitch Class Profile
HSG6D	6-D Vektor von Harte, Sandler und Gasser
Hz	Hertz
i, j	Laufvariablen
K	Anzahl der zu suchenden Nachbarn
K	Anzahl der Klassen
k_{cq}	Tonnummer im Tonvektor
kHz	Kilohertz
KK	Krumhansl/Kessler
KNN	K Nearest Neighbours
$L(\cdot)$	Verlustfunktion
λ	Parameterzusammenfassung eines HMM
\log	Dekadischer Logarithmus
LPC	Late Positive Component
M	Menge
M_p	Die Menge der Realtonigkeiten eines Tonraumes
M	Anzahl der Dimensionen des Beobachtungsvektors (Merkmalsvektors)
m	Grundindex einer Hilfstonigkeit
\mathbf{m}	Beobachtungsvektor (Merkmalsvektor)
MIDI	Musical Instrument Digital Interface
MIR	Music Information Retrieval
MIREX	Music Information Retrieval Exchange

MMN	Mismatch Negativity
<i>ms</i>	Millisekunde
<i>N</i>	Anzahl der Zustände, Anzahl der Klassen
<i>N</i>	Anzahl der Trainingspunkte
<i>n</i>	Erweiterter Index einer Hilfstonigkeit
n-D	n-Dimensional
NN	Neuronale Netzwerke
NNMF	Non negative matrix factorization
O	Beobachtungsmatrix
\mathbf{o}_t	Beobachtungsvektor (Merkmalsvektor)
π	Kreiszahl pi
π	Anfangsverteilung bei Hidden Markov Models
P	Parallelklang
<i>P</i>	Tonartprofil
$P(X)$	Wahrscheinlichkeitsfunktion, wenn X diskret ist
$p(X)$	Wahrscheinlichkeitsdichtefunktion, wenn X kontinuierlich ist
$P(X Y)$	Bedingte Wahrscheinlichkeit von X gegeben Y
PCA	Principal Component Analysis
PCM	Pulse Code Modulation
$\vec{\Psi}$	Summenvektor im kreisförmigen Tonraum
q_t	Betrachteter Zustand
r	Klassenvektor
RATN	Right Anterior-Temporal Negativity
ρ	Realtonigkeit
S	Subdominante
SML	Symmetriemodell
SVM	Support Vector Machines
T	Tonartübergreifender Terzkreis (Key Spanning Circle of Thirds)
T	Tonika
<i>T</i>	Anzahl der Zeitpunkte
θ	Parameter der Hypothesen
t_F	Zeitfensterlänge
<i>TP</i>	Trainingspunkte pro Zustand
TR	Tonartbezogener Terzkreis (Key Related Circle of Thirds)
TY	Temperley
V	Variantklang

V^k	Volumen einer Hypersphäre durch Anzahl an k Nachbarn bestimmt
VSYM	Merkmalsvektor des Symmetriemodells
$w[n, k_{cq}]$	Filterbeschreibung der Constant-Q-Filter
\mathbf{x}	Eigabedatenvektor
\mathbf{X}	Übungsmenge
$X[k_{cq}]$	Tonvektor
ξ	Hilfstonigkeit

Anhang A

Längen der Summenvektoren von Intervallen

Abbildung A.1 stellt die Längen der Summenvektoren aller Intervalle in den Teilräumen F , FR , TR , DR und CR gegenüber. Zunächst ist zu sehen, dass das *Intervall der Prime* (und damit auch der *Oktave*) in allen Teilräumen den längsten Summenvektor verursacht. Das liegt daran, dass beide eine Prime bildenden Tonigkeiten identisch sind und damit auch die resultierenden Tonigkeitsvektoren. Diese im Prinzip triviale Eigenschaft von Tonräumen ist von daher in der Analyse von Musiksignalen zu berücksichtigen, dass das Intervall der Prime auf der einen Seite für die harmonische Entwicklung eines Musikstückes eine untergeordnete Rolle spielt, auf der anderen Seite jedoch stark in Tonräumen hervortritt. Dieses Problem kommt vor allem dann zum Vorschein, wenn die harmonische Entwicklung eines Musikstückes durch Aufsummierung zeitlich benachbarter Tonigkeitskombinationen analysiert werden soll. Eine mögliche Lösung in diesem Zusammenhang besteht darin, lediglich die Differenz zweier zeitlich benachbarter Tonigkeitskombinationen in die Berechnung einzubeziehen.

Das *Intervall der kleinen und großen Sekunde* (Abbildung A.1), welches vor allem für die melodische Entwicklung eines Musikstückes bedeutsam ist, wird hauptsächlich vom tonartbezogenen diatonischen Tonleiterkreis DR hervorgehoben. Hier führt die Sekunde zu einem langen Summenvektor mit einer Länge deutlich größer 1,5. Im tonartbezogenen Terzkreis TR wird die Sekunde stark unterdrückt und führt zu Summenvektoren, die deutlich kleiner als 1,0 und damit deutlich kürzer als die Einzelvektoren der Teiltonigkeiten sind. In den Kreisen F und FR wird das Intervall der Sekunde nur mittelmäßig repräsentiert. In F ist die Länge des Summenvektors nur etwas länger als die von Einzeltonigkeitsvektoren. Im tonartübergreifenden Quintenzirkel F wird die kleine Sekunde (5 Quinten Abstand) relativ stark unterdrückt, während die große Sekunde (2 Quinten)

deutlich hervorgehoben wird.

Die *kleine und die große Terz*, welche die Grundlage von Dur- und Mollakkorden bilden, wird am stärksten vom tonartbezogenen Terzkreis *TR* hervorgehoben. Hier ergeben sowohl die große als auch die kleine Terz Summenvektoren deutlich größer 1,5. Im tonartbezogenen Quintenzirkel *FR* dagegen ist die Terz dasjenige Intervall, welches am stärksten unterdrückt wird (Summenvektorlängen deutlich kleiner 1.0). Im tonartübergreifenden Quintenzirkel *F* sowie im tonartbezogenen diatonischen Tonleiterkreis *DR* wird die Terz weder stark hervorgehoben noch stark unterdrückt (Summenvektorlängen zwischen 1.0 und 1.41).

Der *Tritonus* ist das in den einzelnen Kreisen am gegensätzlichsten dargestellte Intervall. Die beiden den Tritonus bildenden Tonigkeiten liegen in den Kreisen *DR*, *CR* und *F* direkt gegenüber, was zu einer Auslöschung der zugehörigen Tonigkeitsvektoren führt.

Das Intervall der *Quarte bzw. Quinte* führt wie zu erwarten in den Kreisen *F* und *FR* zu sehr langen Summenvektoren. Im Kreis *DR* dagegen kommt es fast zu einer Auslöschung der Teiltonigkeiten und damit zu kurzen Summenvektoren. Im *TR* führt die Quinte zu einem Summenvektor, der ein wenig länger als der Summenvektor einer Einzeltonigkeit ist.

Längen der Summenvektoren von Intervallen

Intervalle	Noten- beispiel	Summenvektorenlängen			
		F	FR	TR	DR/CR
Prime	c-c	2,00	2,00	2,00	2,00
Kleine Sekunde/ Große Septime	e-f	0,52	1,32	0,26	1,93
Gr. Sekunde/ Kleine Septime	c-d	1,73	1,22	0,52	1,73
Kleine Terz/ Große Sexte	a-c	1,41	0,39	1,85	1,41
Große Terz/ Kleine Sexte	c-e	1,00	0,52	1,73	1,00
Tritonus	h-f	0,00	1,85	1,41	0,00
Quarte/ Quinte	c-g	1,93	1,79	1,22	0,52

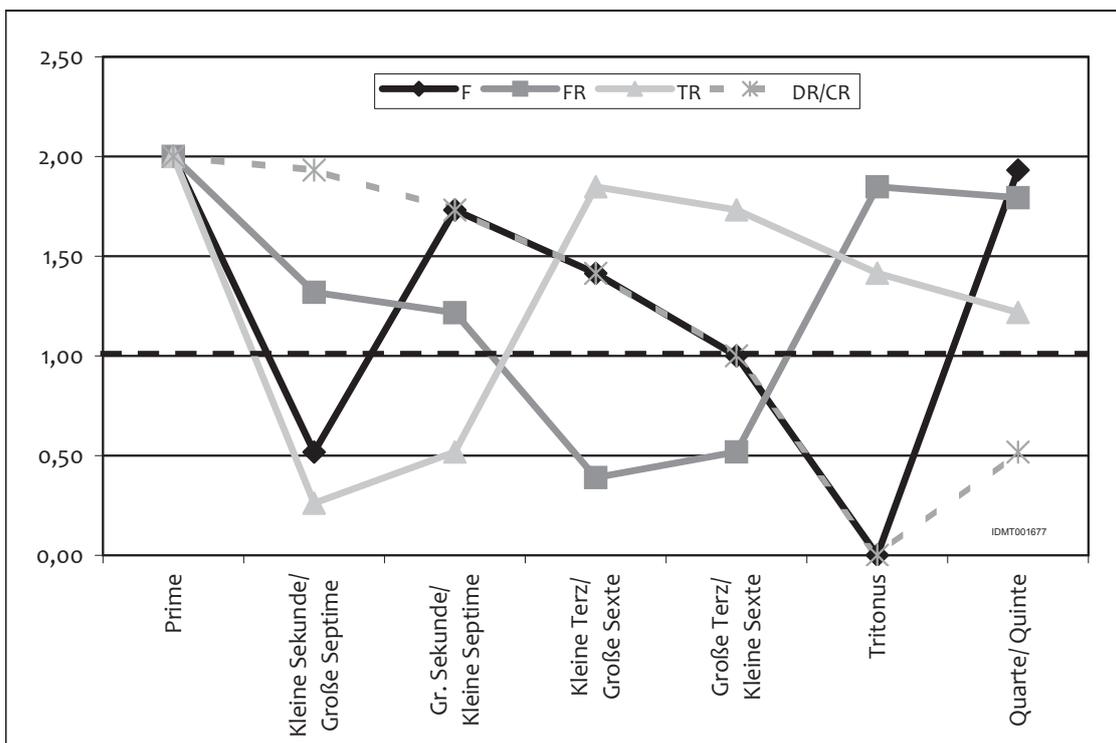


Abbildung A.1 – Die Längen der Summenvektoren von Intervallen in den Teilräumen des Symmetriemodells

Vergleich von Krumhansls C-Dur-Profilen mit den Distanzen im Symmetriemodell

		d	f	a	c	e	g	h
Krumhansl	TTP original, C-Dur	3,48	4,09	3,66	6,35	4,38	5,19	2,88
	P_{KHL} : TTP normiert	0,17	0,35	0,22	1,00	0,43	0,67	0,00
Symmetrie- modell	Distanz zu "c" im F in °	60	30	90	0	120	30	150
	Distanz zu "c" im FR in °	105,00	52,50	157,50	0,00	150,00	52,50	97,50
	Distanz zu "c" im TR in °	150,00	105,00	45,00	0,00	60,00	105,00	165,00
	Distanz zu "c" im DR in °	60,00	150,00	90,00	0,00	120,00	150,00	30,00
	Distanz zu "c" im F normiert	0,40	0,20	0,60	0,00	0,80	0,20	1,00
	Distanz zu "c" im FR normiert	0,67	0,33	1,00	0,00	0,95	0,33	0,62
	Distanz zu "c" im TR normiert	0,91	0,64	0,27	0,00	0,36	0,64	1,00
	Distanz zu "c" im DR normiert	0,40	1,00	0,60	0,00	0,80	1,00	0,20
	P_F : Nähe zu "c" im F	0,60	0,80	0,40	1,00	0,20	0,80	0,00
	P_{FR} : Nähe zu "c" im FR	0,33	0,67	0,00	1,00	0,05	0,67	0,38
P_{TR} : Nähe zu "c" im TR	0,09	0,36	0,73	1,00	0,64	0,36	0,00	
P_{DR} : Nähe zu "c" im DR	0,60	0,00	0,40	1,00	0,20	0,00	0,80	
Gewichtungsfaktoren:		$g_F = 0,00$		$g_{FR} = 1,00$		$g_{TR} = 1,00$		
		$g_{SDR} = 0,00$						
$P_{ALL} = P_F g_F + P_{FR} g_{FR} + \dots + P_{DR} g_{DR}$		0,42	1,03	0,73	2,00	0,68	1,03	0,38
P_{ALL} normiert		0,03	0,40	0,21	1,00	0,19	0,40	0,00
Korre- lationen	Korrelation P_{KHL}, P_F	0,76						
	Korrelation P_{KHL}, P_{FR}	0,68						
	Korrelation P_{KHL}, P_{TR}	0,74						
	Korrelation P_{KHL}, P_{DR}	0,08						
	Korrelation P_{KHL}, P_{ALL}	0,93						

Tabelle A.1 – Vergleich der im Symmetriemodell berechneten Stabilität von Tönen in der Tonart C-Dur mit den von Carol L. Krumhansl gefundenen C-Dur-Profilen (Abbildung 2.5)

Vergleich von Krumhansls a-Moll-Profilen mit den Distanzen im Symmetriemodell

		d	f	a	c	e	g	h	
Krumhansl	Probe Tone Profiles Original für a-Moll	3,53	3,98	6,33	5,38	4,75	3,17	3,52	
	P _K : Probe Tone Profiles genormt	0,11	0,26	1,00	0,70	0,50	0,00	0,11	
Symmetrie- modell	Distanz zu "a" im F in °	30,00	120,00	0,00	90,00	30,00	60,00	60,00	
	Distanz zu "a" im FR in °	52,50	150,00	0,00	157,50	52,50	105,00	105,00	
	Distanz zu "a" im TR in °	105,00	60,00	0,00	60,00	105,00	150,00	150,00	
	Distanz zu "a" im DR in °	150,00	120,00	0,00	90,00	150,00	60,00	60,00	
	Distanz zu "a" im F normiert	0,25	1,00	0,00	0,75	0,25	0,50	0,50	
	Distanz zu "a" im FR normiert	0,33	0,95	0,00	1,00	0,33	0,67	0,67	
	Distanz zu "a" im TR normiert	0,70	0,40	0,00	0,40	0,70	1,00	1,00	
	Distanz zu "a" im DR normiert	1,00	0,80	0,00	0,60	1,00	0,40	0,40	
	P _F : Nähe zu "a" im F normiert	0,75	0,00	1,00	0,25	0,75	0,50	0,50	
	P _{FR} : Nähe zu "a" im FR normiert	0,67	0,05	1,00	0,00	0,67	0,33	0,33	
	P _{TR} : Nähe zu "a" im TR normiert	0,30	0,60	1,00	0,60	0,30	0,00	0,00	
	P _{DR} : Nähe zu "a" im TR normiert	0,00	0,20	1,00	0,40	0,00	0,60	0,60	
	Gewichtungsfaktoren:		g _F = 0,00	g _{FR} = 0,00	g _{TR} = 1,00				
			g _{SDR} = 0,00	g _{SYM} = 0,00					
	P _{ALL} = P _F g _F + P _{FR} g _{FR} + ... + P _{DR} g _{DR}	0,30	0,60	1,00	0,60	0,30	0,00	0,00	
	P _{ALL} normiert	0,30	0,60	1,00	0,60	0,30	0,00	0,00	
Korre- lationen	Korrelation P _K , P _F	0,34							
	Korrelation P _K , P _{FR}	0,37							
	Korrelation P _K , P _{TR}	0,87							
	Korrelation P _K , P _{DR}	0,41							
	Korrelation P _K , P _{ALL}	0,87							

Tabelle A.2 – Vergleich der im Symmetriemodell berechneten Stabilität von Tönen in der Tonart a-Moll mit den von Carol L. Krumhansl gefundenen a-Moll-Profilen (Abbildung 2.5)

Anhang B

Beispiele zur Anwendung des SYM-Operators

B.1 Einordnung einer chromatischen Tonigkeit in die Quintreihe

Wir haben einen Chromavektor gegeben, welcher 12 Energiewerte für jede chromatische Tonigkeit liefert. Um die Summenvektoren in den kreisförmigen Tonräumen des Symmetriemodells zu berechnen, ist es zunächst erforderlich, die 12 chromatischen Tonigkeiten an die richtige Stelle in der Quintreihe einzuordnen. Unter Annahme einer Zuordnung von Tonigkeiten zu Indizes n_C wie in Tabelle 4.1 beschrieben kann eine Tonigkeit n_C in die Quintreihe wie folgt eingeordnet werden:

$$n_F = \text{SYM3}(49n_C, 84, 7n_k) \quad (\text{B.1})$$

Das Produkt $49n_C$ bildet die Tonigkeit n_C zunächst an eine der Stellen in der Quintreihe ab, welche die Tonigkeit n_C besitzt. Die Tonigkeit *dis* ($n_C = 1$) wird z.B. an die Stelle $n_F = 49$ abgebildet, das ist die Tonigkeit *dis*. Angenommen die Tonart wäre Es-Dur ($n_k = -3$), so müsste die Tonigkeit *dis* nicht an der Stelle 49, sondern als *es* an der Stelle $n_F = -35$ erscheinen. Genau diese Unterscheidung wird durch den SYM3-Operator vorgenommen. Dieser ordnet die Tonigkeit n_C in einen Bereich von +6 Quinten (84 Halbtöne) um den Symmetrieton ($7n_k$) eines gewählten diatonischen Tonssets ein¹.

¹ $n_F = \text{SYM}(1*49, 84, 7*-3) = -35$

B.2 Test, ob eine Tonigkeit Bestandteil eines diatonischen Tonsets ist

Um eine Tonigkeit in einen der tonartbezogenen Räume *FR*, *TR* oder *DR* einordnen zu können, muss zunächst entschieden werden, ob diese Tonigkeit Bestandteil des diatonischen Tonsets ist, welche vom gewählten tonartbezogenen Modell repräsentiert wird. Dazu wird zunächst mit Hilfe von Gleichung B.1 die tonartabhängige Position der Tonigkeit in der Quintreihe n_F berechnet. Für diatonische Tonigkeiten gilt dann:

$$-21 \leq n_F - 7n_k < +21 \quad (\text{B.2})$$

Dieser Gleichung liegt die Annahme zugrunde, dass sich diatonische Tonigkeiten in einem Bereich $+3$ Quinten (21 Halbtöne) um die Symmetrietonigkeit ($n_F = 7n_k$) des diatonischen Tonsets befinden.

B.3 Einordnung einer Tonigkeit in einen der Unterräume des Symmetriemodells

Wir sind nun in der Lage, eine chromatische Tonigkeit n_C korrekt in die Quintreihe einzuordnen. Im nächsten Schritt soll die Tonigkeit n_C nun in die Räume *FR*, *TR* und *DR* eingeordnet werden. Dazu wird zunächst mit Hilfe von Gleichung B.2 überprüft, ob die Tonigkeit n_C zum diatonischen Tonset n_K gehört. Anschließend wird mit Hilfe von Gleichung B.1 der Index n_F berechnet. Nun folgt die Einordnung in die tonartbezogenen Kreise *FR*, *TR* und *DR*:

$$n_{FR} = \text{SYM}(n_F - 7n_k, 48) \quad (\text{B.3})$$

$$n_{TR} = \text{SYM}(n_F - 7n_k - 12, 24) \quad (\text{B.4})$$

$$n_{DR} = \text{SYM}(n_F - 7n_k, 12) \quad (\text{B.5})$$

Die Konstante $-7n_k$ dreht die tonartbezogenen Kreise so, dass die Symmetrietonigkeit des diatonischen Tonsets n_k den Wert 0 bekommt. Im Fall des TRs wird noch eine Konstante -12 addiert, welche den *TR* so dreht, dass der Symmetrieton unten erscheint (aufgrund existierender Konventionen).

B.4 Eigenschaften einer Tonigkeit in einem gegebenen diatonischen Tonset

Mit Hilfe der vorherigen Gleichungen können wir nun eine beliebige Tonigkeit n_C in die Unterräume des Symmetriemodells einordnen. Indem wir den SYM-Operator anwenden, ist es nun möglich, verschiedene musikalische Eigenschaften von Tonigkeiten einfach zu bestimmen. Es gilt: Eine Tonigkeit ist Grundton eines Durakkordes, wenn gilt: $-24 < n_{FR} < 0$, eines Mollakkordes, wenn gilt: $0 \leq n_{FR} \leq 21$, und tiefster Ton eines verminderten Dreiklanges, wenn gilt: $n_{FR} = 21$. Weiterhin kann einfach berechnet werden, ob die Tonika Bestandteil des Tonika, Dominant- oder Subdominantbereichs (Abschnitt 5.3.2) ist. Es gilt: Eine Tonigkeit ist Bestandteil des Subdominantbereiches, wenn gilt: $-12 \leq n_{TR} \leq 0$, des Tonikabereiches, wenn gilt: $-6 \leq n_{TR} \leq 6$, und des Dominantbereiches, wenn gilt: $0 \leq n_{TR} \leq +12$.

B.5 Vorhersage des Auflösungsbestrebens einer Tonigkeit in einer gegebenen Tonart

Nach Abschnitt 5.5.4 kann die Stabilität einer Tonigkeit t in einer gegebenen Tonart n_K in Form der Winkeldistanz zwischen Tonartgrundton gn und der Tonigkeit tt im TR angenähert werden. Die Berechnung der Stabilität s ergibt sich damit wie folgt:

$$s = (n_{TR_{gt}} - n_{TR_{tt}}) \bmod 24 \quad (\text{B.6})$$

24 ist die Halbtonauflösung des TRs (g_{TR}). Die Indizes $n_{TR_{gt}}$ und $n_{TR_{tt}}$ können mit Hilfe der zuvor beschriebenen Gleichungen berechnet werden.

B.6 Berechnung der Tonigkeitsbezeichnung einer Tonigkeit

Nehmen wir an, wir haben zwei Tonigkeiten $n_F = -35$ und $n_F = +49$. Um die musiktheoretisch korrekte Bezeichnung zu ermitteln, muss zunächst der Stammtton $n_{F_{st}}$ der Tonigkeiten ermittelt werden. Der Stammtton ist einer der sieben Tonigkeiten f, c, g, d, a, e und h und ergibt sich folgendermaßen:

$$n_{F_{st}} = \text{SYM}(n_F, 49) \quad (\text{B.7})$$

Anschließend wird ermittelt, wie oft der Stammtton erhöht oder erniedrigt wurde, d.h. die Anzahl der Vorzeichen n_V (Kreuze oder Bs) wird ermittelt. Diese ergibt sich aus

der Anzahl der Windungen in der Stammttonspirale (Gleichung B.9a), welche zwischen Stammtton n_{ST} und n_F liegen. Es gilt:

$$n_V = (n_F - n_{Fst})/49 \quad (\text{B.8})$$

Für unsere Tonigkeit $n_F = -35$ erhalten wir $n_{Fst} = 14$. Dies ist entsprechend Abbildung 4.3 die Tonigkeit e. Als Anzahl der Vorzeichen erhalten wir -1, was einer einfachen Erniedrigung (b) entspricht. Die Tonigkeit mit dem Index n_F hat also die Bezeichnung e_b bzw. es. Für die Tonigkeit $n_F = 49$ erhalten wir $n_{Fst} = 0$ und $n_V = +1$. Dies entspricht der Tonigkeit $d_\#$ bzw. dis.

B.7 Finden der einfachsten enharmonisch equivalenten Tonart

Mit Hilfe des *SYM*-Operators ist es möglich, zu einem diatonischen Tonset mit beliebig vielen Vorzeichen die zugehörige Tonart zwischen *Ges*-Dur und *Fis*-Dur zu finden. Gegeben ist ein diatonisches Tonset mit n_k Vorzeichen. Gesucht ist das enharmonisch identische Tonset $n_{k_{simple}}$ zwischen *Fis*-Dur und *Ges*-Dur. Es gilt:

$$n_{k_{simple}} = \text{SYM}(n_k, 12) \quad (\text{B.9})$$

Setzt man z.B. die Tonart *Ais*-Dur (10 Kreuze, $n_k=10$) in Gleichung B.9 ein, so erhält man die Tonart *B*-Dur (2Bs, $n_k=-2$): $\text{SYM}(+10, 12) = -2$.

Anhang C

Funktionssymbole

Da die Zuordnung von Funktionen zu Akkorden ein starkes Hilfsmittel für die Harmonieanalyse darstellt, wurde ein Symbolsystem entwickelt. Hierbei werden die Funktionen Tonika, Subdominante und Dominante über die Symbole *T*, *S* und *D* für Durtonarten bzw. *t*, *s* und *d* für reine Molltonarten abgekürzt. Die Nebenfunktionen werden durch Anhängen der Buchstaben *g* oder *p* bzw. *G* oder *P* (*p/P* - Parallelklang, *g/G* - Gegenklang) an das Symbol der Hauptfunktion abgekürzt. Handelt es sich beim Nebendreiklang um einen Durakkord, dann werden Großbuchstaben angehängt, handelt es sich um einen Mollakkord, dann werden Kleinbuchstaben hinzugefügt¹. Das Symbol *Tp* bezeichnet so z.B. die Mollparallele zur Durtonika. In Abbildung C.1 werden die Funktionssymbole in die Terzspirale eingeordnet. Ein Bezug zu Akkordsymbolen kann durch Vergleich mit Abbildung 5.10 hergestellt werden. Eine Gesamtübersicht über die wichtigsten Funktionssymbole kann im Anhang von [83] gefunden werden.

¹ Ein Beispiel: Tonika *T*: C-Dur; Tonikaparallele *Tp*: a-Moll; Tonikagegenklang *Tg*: e-Moll;

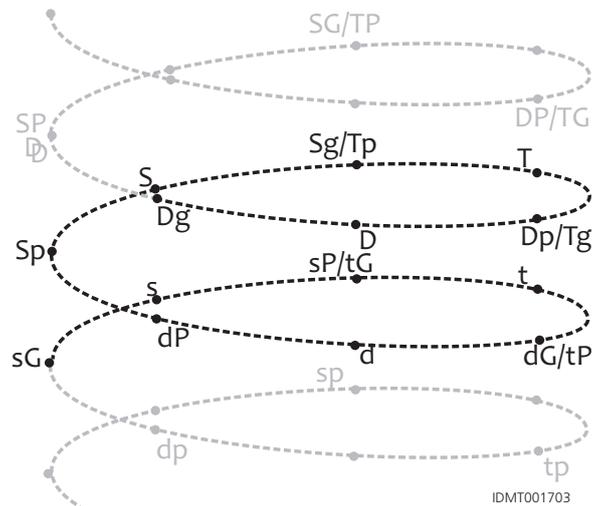


Abbildung C.1 – Einordnung der Funktionssymbole in die Terzspirale

Anhang D

Mathematische Grundlagen

D.1 Statistische Auswertung von Ergebnislisten

Ein Analysealgorithmus wird in der Regel mit mehreren zu untersuchenden Objekten (in dieser Arbeit: Musikstücke) getestet, d.h. das Ergebnis der Evaluierung ist eine Ergebnisliste, die aussagt, wie oft der Algorithmus das richtige Ergebnis ermitteln konnte oder wie hoch die Erkennungsrate für jedes Objekt war. Bei der Akkorderkennung wird z.B. eine Trefferrate an richtig erkannten Akkorden oder richtig zugeordneten Zeitfenstern für jedes Musikstück angegeben. Letztendlich wird ein einziger Mittelwert über die gesamte Ergebnisliste berechnet, der die Leistungsfähigkeit des Algorithmus widerspiegelt. Wenn zwei Algorithmen verglichen werden sollen, dann gibt es für jeden einen Mittelwert ihrer Ergebnislisten. Um aussagen zu können, welcher der beiden Algorithmen besser ist, genügt es allerdings nicht, einfach die beiden Mittelwerte zu vergleichen, sondern es ist herauszufinden, ob die beiden Mittelwerte sich signifikant unterscheiden. Für diese Aufgabe gibt es den sogenannten „Friedman-Test“.¹

Friedman-Test

Der Friedman-Test ist ein parameterfreier Test und wird zum Vergleich mehrerer ($c \geq 2$) abhängiger Stichproben (Ergebnislisten) verwendet. Um den Test durchführen zu können, müssen folgende Voraussetzungen erfüllt sein:

1. Das zu untersuchende Merkmal muss stetig sein.
2. Es muss eine Rangliste ermittelt werden können.

¹ Der Test wurde nach dem amerikanischen Mathematiker Milton Friedman benannt, der ihn 1937 entwickelt hat.

3. Die Stichproben müssen abhängig sein.

Die Ergebnislisten mit den Einzelergebnissen X werden zu einer Ergebnismatrix zusammengefasst, wobei die Ergebnisse einer Zeile zum gleichen Objekt (Musikstück) gehören müssen. Tabelle D.1 zeigt die Ergebnismatrix.

		Algorithmus Nr.				
		1	2	.	.	c
Musikstück	1	X_{11}	X_{12}	.	.	X_{1c}
	2	X_{21}	X_{22}	.	.	X_{2c}

	n	X_{n1}	X_{n2}	.	.	X_{nc}

Tabelle D.1 – Ergebnismatrix - Zusammenfassen mehrerer Ergebnislisten

Innerhalb jeder Zeile wird eine Rangzuordnung vorgenommen, d.h. es gibt genau einen Wert mit Rang 1, einen mit Rang 2 usw. Die Summe in jeder Zeile ist damit genau $1/2 \cdot c \cdot (c + 1)$. Tabelle D.2 zeigt die Rangmatrix mit den Rangzuordnungen.

		Algorithmus Nr.					Zeilensumme
		1	2	.	.	c	
Musikstück	1	R_{11}	R_{12}	.	.	R_{1c}	$1/2 \cdot c \cdot (c + 1)$
	2	R_{21}	R_{22}	.	.	R_{2c}	$1/2 \cdot c \cdot (c + 1)$

	n	R_{n1}	R_{n2}	.	.	R_{nc}	$1/2 \cdot c \cdot (c + 1)$
Spaltensumme		R_1	R_2	.	.	R_c	$n \cdot 1/2 \cdot c \cdot (c + 1)$

Tabelle D.2 – Rangmatrix

Wenn die Stichproben nur zufällig voneinander abweichen (Nullhypothese H_0) und damit die Mittelwerte der Ergebnislisten, dann müsste in jeder Stichprobe jeder Rang (1, ... ,k) in etwa gleich häufig vorkommen und die Summe jeder Spalte der Rang-Ergebnismatrix gleich groß sein. Die Spaltensumme ist:

$$R_j = \sum_{i=1}^n R_{ij}, \text{ mit } j = 1, \dots, c \quad (\text{D.1})$$

Der Erwartungswert der Nullhypothese für alle R_j ergibt sich zu:

$$E(R_j) = \frac{n(c+1)}{2} \forall j = 1, \dots, c \quad (\text{D.2})$$

Friedman schlägt die folgende Größe F als Teststatistik vor:

$$F = \frac{12}{n \cdot c \cdot (c+1)} \sum_{j=1}^c [R_j - E(R_j)] \quad (\text{D.3})$$

F ist asymptotisch χ^2 -verteilt mit $(c-1)$ Freiheitsgraden. Das Ergebnis F ist gegen die Verteilung auf Annahme der Nullhypothese zu testen. Bei Ablehnung der Nullhypothese sind mindestens zwei Algorithmen signifikant unterschiedlich.

Da aber nicht nur herausgefunden werden soll, ob es einen signifikanten Unterschied zwischen den Ergebnissen gibt, sondern auch für welche Stichproben dies zutrifft, muss ein mehrfacher paarweiser Test aller Stichproben durchgeführt werden. Diese Berechnung erfolgt nach dem Friedman-Rangsummentest. Eine Einführung dazu ist in [61] S. 622 ff. zu finden.

D.2 Distanzmaße

Zur Berechnung einer Distanz $d(x,y)$ zwischen zwei Punkten x und y in einem Raum können verschiedene Metriken verwendet werden. In dieser Arbeit kommen die *euklidische Distanz*, die *Kosinus-Distanz* und die *Manhattan-Distanz* zum Einsatz.

Euklidische Distanz

Die *euklidische Distanz* ist wie folgt definiert:

$$d_{\text{euklidisch}}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (\text{D.4})$$

Kosinus-Distanz

Der *Kosinus Abstand* berechnet den Kosinus des eingeschlossenen Winkels zwischen den beiden gerade betrachteten Punkten im Raum und bildet mit ihm die Differenz zu $,1'$. Er ist wie folgt definiert:

$$d_{\text{Kosinus}}(x, y) = 1 - \cos(\angle(x, y)) \quad (\text{D.5})$$

Manhattan-Distanz

Die *Manhattan-Distanz* berechnet die Abstände der einzelnen Dimensionen unabhängig voneinander und bildet daraus die Gesamtdistanz:

$$d_{\text{Manhattan}}(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (\text{D.6})$$

Anhang E

Tonarterkennung: Verwendete Teststücke und deren Tonarten

Beatles

Nr.	Titel	Tonset	Tonika
1	I saw her standing there	E	E
2	Misery	C	C
3	Anna (go to him)	D	D
4	Chains	B	B
5	Boys	E	E
6	Ask me why	E	E
7	Please please me	E	E
8	Love me do	G	G
9	P.s. I love you	D	D
10	Baby it's you	G	G
11	Do you want to know a secret	E	E
12	A taste of honey	E	fis
13	There's a place	E	E
14	Twist and shout	D	D
15	It won't be long	E	E
16	All I've got to do	E	cis
17	All my loving	E	E
18	Don't bother me	D	e
19	Little child	E	E
20	Till there was you	F	F
21	Please mister postman	A	A
22	Roll over beethoven	D	D
23	Hold me tight	F	F
24	You really got a hold on me	A	A
25	I wanna be your man	E	E
26	Devil in her heart	G	G
27	Not a second time	G	G
28	Money	E	E
29	A hard day's night	C	G
30	I should have known better	G	G
31	If I fell	D	D
32	I'm happy just to dance with	E	E
33	And I love her	E	fis
34	Tell me why	D	D
35	Can't buy me love	C	C
36	Any time at all	D	D
37	I'll cry instead	G	G
38	Things we said today	C	a
39	When I get home	C	C
40	You can't do that	G	G
41	I'll be back	C	a
42	No reply	C	C
43	I'm a loser	G	G
44	Baby's in black	A	A
45	Rock and roll music	A	A
46	I'll follow the sun	C	C
47	Mr moonlight	Fis	Fis
48	Kansas city, hey,hey,hey,hey	G	G
49	Eight days a week	D	D
50	Words of love	A	A

Tabelle E.1 – Beatles-Stücke und deren Tonarten – Teil 1

51	Honey don't	E	E
52	Every little thing	A	A
53	I don't want to spoil the party	G	G
54	What you're doing	D	D
55	Everybody's trying to be my baby	E	E
56	Help	A	A
57	The night before	D	D
58	You've got to hide your love	C	G
59	I need you	A	A
60	Another girl	A	A
61	You're going to lose that gir	E	E
62	Ticket to ride	A	A
63	Act naturally	G	G
64	It's only love	C	C
65	You like me too much	G	G
66	Tell me what you see	G	G
67	I've just seen a face	A	A
68	Yesterday	F	F
69	Dizzy miss lizzy	A	A
70	Drive my car	D	D
71	Norwegian wood	E	E
72	You won't see me	A	A
73	Nowhere man	E	E
74	Think for yourself	G	G
75	The word	D	D
76	Michelle	F	F
77	What goes on	E	E
78	Girl	Es	c
79	I'm looking through you	As	As
80	In my life	A	A
81	Wait	A	fis
82	If i needed someone	D	A
83	Run for your life	D	D
84	Taxman	G	D
85	Eleanor rigby	G	e
86	I'm only sleeping	Ges	es
87	Love you to	F	C
88	Here, there and everywhere	G	G
89	Yellow submarine	Fis	Fis
90	She said she said	Es	B
91	Good day sunshine	A	A
92	And your bird can sing	E	E
93	For no one	H	H
94	Doctor robert	A	A
95	I want to tell you	A	A
96	Got to get you into my life	G	G
97	Tomorrow never knows	F	C
98	Sgt.pepper's lonely hearts club band	G	G
99	With a little help from my friends	E	E
100	Lucy in the sky with diamonds	A	A

Graue Felder sind Dorisch, Phrygisch, Lydisch oder Mixolydisch.

Tabelle E.2 – Beatles-Stücke und deren Tonarten – Teil 2

RockPop

Nr.	Titel	Tonset	Tonika
1	ACDC - COD	C	G
2	ACDC - Highway to hell	D	A
3	ACDC - TNT	Es	Es
4	ACDC - Touch too much	G	e
5	Aerosmith - Janie's got a gun	F	F
6	Cheap Trick - I want you to want me	A	A
7	Die Fantastischen 4 - MfG	F	g
8	Die Happy - Supersonic speed	G	e
9	Die Sterne - Was hat dich blos so ruiniert	G	a
10	Eagles - I can't tell you why	D	D
11	Echt - Alles wird sich ändern	D	D
12	Echt - Du trägst keine Liebe in Dir	B	B
13	Echt - Weinst Du	C	e
14	Echt - Wo bist Du jetzt	H	H
15	Eminem - Toy soldiers	E	E
16	Eric Carmen - All by myself	F	F
17	Freundeskreis - Anna	G	e
18	Good Charlotte - I just wanna live	Des	b
19	Hanson - Mmmmbop	A	A
20	Hootie & The Blowfish - Drowning	C	G
21	Hootie & The Blowfish - Goodbye	G	G
22	Hootie & The Blowfish - Hannah Jane	G	G
23	Hootie & The Blowfish - Hold my hand	H	H
24	Hootie & The Blowfish - I'm goin' home	E	E
25	Hootie & The Blowfish - Let her cry	G	G
26	Hootie & The Blowfish - Look away	G	G
27	Hootie & The Blowfish - Not even the trees	D	D
28	Hootie & The Blowfish - Only wanna be with you	Fis	Fis
29	Hootie & The Blowfish - Running from an angel	E	E
30	Hootie & The Blowfish - Time	D	D
31	INXS - Mystify	C	d
32	Jam & Spoon - Be Angeled	B	F
33	Joan Osborne - One of us	A	fis
34	Michael Jackson - Heal the world	A	A
35	Moby - Bodyrock	Fis	Fis
36	Moby - Everloving	G	e
37	Moby - Find my baby	Es	g
38	Moby - Guitar flute and strings	G	e
39	Moby - Honey	Des	b
40	Moby - If things were perfect	C	a
41	Moby - Inside	A	A
42	Moby - Machete	A	fis
43	Moby - My weakness	E	E
44	Moby - Natural blues	Des	b
45	Moby - Porcelain	Es	g
46	Moby - Rushing	G	D
47	Moby - South side	E	cis
48	Moby - Why does my heart feels so bad	G	a
49	Pet Shop Boys - West end girls	A	E
50	Pet Shop Boys - Where the streets have no name	B	B

Tabelle E.3 – RockPop-Stücke und deren Tonarten – Teil 1

51	Petula Clark - Downtown	E	E
52	REM - Drive	C	d
53	REM - Everybody hurts	D	D
54	REM - Find the river	D	D
55	REM - Ignoreland	C	d
56	REM - Losing my religion	C	a
57	REM - Man on the moon	G	C
58	REM - Monty got a raw deal	B	g
59	REM - New Orleans instrumental No 1	C	C
60	REM - Nightswimming	G	G
61	REM - Star me kitten	G	G
62	REM - Sweetness follows	D	D
63	REM - The sidewinder sleeps tonight	C	C
64	REM - Try not to breathe	G	D
65	Republica - Ready to go	G	D
66	Selig - Ohne Dich	C	C
67	Shaggy - Angel	Fis	Fis
68	Smashing Pumpkins - Today	Es	Es
69	Sonny & Cher - I got you babe	F	F
70	Such A Surge - Jetzt ist gut	Fis	dis
71	The Corrs - Breathless	H	H
72	The Cure - Boys dont cry	A	A
73	The Rasmus - In the shadow	A	fis
74	Thomas D - Frisör	Des	Des
75	Travis - Sing	A	fis
76	Udo Lindenberg Live - Bunte Republik Deutschland	C	a
77	Udo Lindenberg Live - Der frische Wind von Berlin	Es	c
78	Udo Lindenberg Live - Die Polizistin	D	D
79	Udo Lindenberg Live - Goodbye Sailor	F	F
80	Udo Lindenberg Live - Hallo DDR	E	E
81	Udo Lindenberg Live - Horizont	C	a
82	Udo Lindenberg Live - Ich lieb dich überhaupt nicht mehr	C	a
83	Udo Lindenberg Live - Mädchen aus Ostberlin	F	d
84	Udo Lindenberg Live - Nana M	A	A
85	Udo Lindenberg Live - Odyssee	E	fis
86	Udo Lindenberg Live - Reeperbahn	Es	c
87	Udo Lindenberg Live - Sonderzug nach Pankow	B	B
88	Udo Lindenberg Live - Turn away	E	E
89	Udo Lindenberg Live - Wenn die Heizer kommen	C	a
90	Van Halen - Jump	C	C

Graue Felder sind Dorisch, Phrygisch, Lydisch oder Mixolydisch.

Tabelle E.4 – RockPop-Stücke und deren Tonarten – Teil 2

Anhang F

Ergebnisse Akkorderkennung

Nr.	Beatles-Lied	Manhattan			Manhattan			Manhattan			Kosinus		
		500			500			500			30		
		3	3	3	3	3	3	15	15	15	15	15	15
		Manhattan			Manhattan			Manhattan			Kosinus		
		Euklidisch			Euklidisch			Euklidisch			Euklidisch		
		Distanzmaß			Distanzmaß			Distanzmaß			Distanzmaß		
		K/TP			K/TP			K/TP			K/TP		
		TP			TP			TP			TP		
		KNN			HMM			HMM/KNN			HMM/KNN		
		CHROMA	HSG6D	VSYM	CHROMA	HSG6D	VSYM	CHROMA	HSG6D	VSYM	CHROMA	HSG6D	VSYM
1	I saw her standing there	55,9	40,9	52,5	59,8	67,0	54,8	85,3	74,1	80,8	85,3	74,1	80,8
2	Misery	56,9	51,2	59,0	66,3	82,3	72,3	87,0	88,3	88,3	87,0	88,3	88,3
3	Anna(go to him)	47,3	39,6	47,2	49,3	58,6	49,4	77,2	67,6	75,4	77,2	67,6	75,4
4	Chains	39,2	37,1	44,8	46,8	60,2	48,5	58,0	56,6	65,6	58,0	56,6	65,6
5	Boys	51,2	30,8	46,9	58,4	42,1	45,3	74,5	56,5	68,2	74,5	56,5	68,2
6	Ask me why	45,1	50,2	50,0	58,8	71,3	61,3	69,9	76,9	75,7	69,9	76,9	75,7
7	Please please me	64,8	51,0	65,5	66,9	63,7	65,7	77,6	71,6	82,9	77,6	71,6	82,9
8	Love me do	37,8	28,0	36,1	31,7	41,7	30,7	52,7	47,1	57,7	52,7	47,1	57,7
9	P.S. I love you	72,4	61,2	70,6	73,5	81,6	72,9	82,2	80,6	88,1	82,2	80,6	88,1
10	Baby it's you	47,5	48,6	52,7	50,6	73,5	58,0	84,7	82,8	89,2	84,7	82,8	89,2
11	Do you want to know a secret	49,3	44,1	53,5	54,0	64,2	59,5	63,5	64,1	70,4	63,5	64,1	70,4
12	A taste of honey	52,0	53,9	55,9	63,0	63,7	57,9	71,1	74,1	70,8	71,1	74,1	70,8
13	There's a place	56,2	38,0	59,9	58,2	58,7	56,7	69,1	49,9	79,5	69,1	49,9	79,5
14	Twist and shout	56,0	42,0	54,2	54,6	38,4	47,8	69,0	56,1	73,6	69,0	56,1	73,6
15	It won't be long	55,3	44,2	49,2	59,7	65,5	50,4	71,0	64,9	65,6	71,0	64,9	65,6
16	All i've got to do	36,5	38,4	43,9	37,6	61,9	48,2	68,8	59,8	74,7	68,8	59,8	74,7
17	All my loving	56,9	51,5	56,7	58,9	65,6	58,4	76,1	71,3	77,9	76,1	71,3	77,9
18	Don't bother me	36,9	35,5	37,0	39,5	49,8	39,6	51,9	57,9	58,9	51,9	57,9	58,9
19	Little child	58,2	42,0	49,4	68,3	57,8	41,3	79,2	61,6	70,7	79,2	61,6	70,7
20	Till there was you	30,2	31,4	34,3	24,3	46,2	38,9	54,4	45,4	47,5	54,4	45,4	47,5
21	Please mister postman	56,3	50,8	51,6	67,8	81,5	55,1	82,5	81,7	84,4	82,5	81,7	84,4
22	Roll over beethoven	44,5	34,7	40,9	47,1	39,4	40,6	65,8	47,8	61,6	65,8	47,8	61,6
23	Hold me tight	28,7	29,3	33,0	26,3	43,1	27,8	49,4	46,8	52,9	49,4	46,8	52,9
24	You really got a hold on me	46,9	44,3	47,9	59,2	70,4	50,5	79,6	73,8	79,4	79,6	73,8	79,4
25	I wanna be your man	58,2	35,7	50,7	70,7	64,6	42,7	82,8	55,9	77,6	82,8	55,9	77,6
26	Devil in her heart	53,1	49,7	56,0	77,0	67,4	62,0	80,7	76,7	80,0	80,7	76,7	80,0
27	Not a second time	26,2	28,2	30,3	30,4	28,1	35,7	46,0	42,4	51,1	46,0	42,4	51,1
28	Money	46,7	35,9	41,5	53,9	54,7	33,5	64,5	57,7	43,8	64,5	57,7	43,8
29	A hard day's night	48,4	37,0	46,4	54,6	36,3	48,8	71,6	51,3	72,7	71,6	51,3	72,7
30	I should have known better	62,8	54,0	60,4	64,2	69,0	64,7	74,6	73,4	76,2	74,6	73,4	76,2
31	If i fell	52,2	49,7	50,4	53,3	67,7	52,2	76,8	75,3	72,9	76,8	75,3	72,9

Tabelle F.1 – Ergebnisse Akkorderkennung – Teil 1

32	I'm happy just to dance with	47,5	43,2	45,7	53,6	56,3	52,7	62,9	61,6	65,4
33	And I love her	49,7	51,6	47,2	60,1	76,4	59,2	79,3	84,5	74,4
34	Tell me why	39,7	34,4	36,4	41,9	43,0	38,5	52,2	49,5	48,0
35	Can't buy me love	42,2	34,8	43,8	55,3	49,8	39,6	79,8	58,7	71,2
36	Any time at all	51,7	45,6	46,3	61,8	63,6	51,9	71,6	64,4	75,0
37	I'll cry instead	66,0	54,6	61,3	67,4	83,7	66,7	89,8	80,7	88,1
38	Things we said today	36,5	38,2	39,5	67,7	46,1	50,9	68,9	75,3	70,9
39	When i get home	45,1	40,8	48,8	54,6	31,2	50,7	64,5	51,5	69,8
40	You can't do that	51,5	33,7	46,1	55,9	34,2	43,2	75,7	51,4	75,1
41	I'll be back	61,0	62,5	62,2	74,3	73,9	71,4	79,2	82,1	86,2
42	No reply	47,5	54,7	60,5	42,5	81,5	68,0	87,2	84,9	87,7
43	I'm a loser	47,7	44,5	51,8	51,7	58,2	58,6	68,2	66,6	69,9
44	Baby's in black	69,5	57,2	67,4	74,8	86,2	67,0	85,9	83,9	84,8
45	Rock and roll music	66,6	50,4	58,5	74,1	84,9	59,2	86,2	86,1	88,7
46	I'll follow the sun	44,7	43,6	66,3	43,6	63,0	72,1	74,1	72,6	81,9
47	Mr moonlight	49,8	46,9	50,0	59,8	81,7	70,9	91,1	90,1	87,3
48	Kansas city, hey,hey,hey,hey	42,8	39,1	43,8	50,9	82,2	43,2	79,3	78,4	78,2
49	Eight days a week	59,0	55,0	60,2	68,5	79,5	66,4	83,4	81,7	86,7
50	Words of love	67,7	57,8	71,2	60,6	75,8	67,2	87,3	84,2	86,0
51	Honey don't	63,4	53,4	66,6	63,2	78,9	64,6	87,1	78,1	85,9
52	Every little thing	65,3	56,9	66,1	70,0	79,6	66,8	78,9	76,5	79,9
53	I don't want to spoil the party	65,2	56,0	64,2	73,2	78,0	68,4	81,2	81,7	83,5
54	What you're doing	65,4	62,7	63,5	68,5	81,2	69,8	86,5	85,2	86,2
55	Everybody's trying to be my baby	66,1	52,1	61,5	73,1	76,3	65,1	85,2	79,9	85,4
56	Help	50,2	51,9	53,4	60,1	84,8	61,2	89,6	85,5	91,6
57	The night before	43,5	42,7	51,1	33,1	62,6	54,7	69,3	66,4	70,2
58	You've got to hide your love	40,4	53,3	64,5	34,9	66,8	63,5	76,1	79,2	80,7
59	I need you	56,3	56,9	64,2	55,5	84,4	68,1	87,8	87,2	87,8
60	Another girl	49,8	44,7	47,2	31,0	62,6	45,4	72,9	66,9	70,5
61	You're going to lose that gir	41,3	46,1	48,6	34,5	59,1	56,4	73,6	71,4	72,7
62	Ticket to ride	55,3	51,2	55,9	40,0	72,1	58,1	82,5	78,5	81,2
63	Act naturally	40,6	54,5	67,4	50,0	65,5	67,6	83,6	84,7	88,9
64	It's only love	14,4	48,0	53,4	17,6	67,7	55,8	77,4	74,6	80,9
65	You like me too much	45,8	54,8	58,4	54,3	77,5	64,5	82,7	82,1	82,9
66	Tell me what you see	25,8	46,8	55,4	9,7	61,9	52,4	70,4	68,5	68,4
67	I've just seen a face	48,5	54,3	59,3	47,6	75,7	63,8	77,6	80,2	82,8

Tabelle F2 – Ergebnisse Akkorderkennung – Teil 2

68	Yesterday	10,7	42,4	52,8	19,9	48,9	51,8	56,2	60,5	67,1
69	Dizzy miss lizzy	44,9	36,5	42,0	36,5	58,3	38,5	81,2	60,9	80,4
70	Drive my car	37,8	38,7	45,0	43,8	57,6	44,8	62,7	57,2	60,9
71	Norwegian wood	44,4	62,9	73,2	79,0	78,7	79,6	85,0	89,4	88,4
72	You won't see me	48,9	50,3	63,8	61,1	72,7	64,8	81,1	76,8	79,1
73	Nowhere man	36,0	41,4	51,1	35,6	64,1	50,6	67,9	64,5	64,7
74	Think for yourself	51,3	49,8	58,6	37,5	59,7	56,2	61,2	67,5	69,5
75	The word	39,9	39,2	50,1	19,6	55,4	41,0	85,1	51,4	64,7
76	Michelle	21,6	31,6	33,5	26,5	38,6	41,1	44,4	57,8	44,6
77	What goes on	39,0	45,9	56,6	45,5	72,6	64,1	82,9	76,4	83,3
78	Girl	17,3	38,8	28,4	41,8	71,0	61,7	55,7	57,7	54,6
79	I'm looking through you	28,5	25,7	21,6	14,8	39,2	49,5	40,6	46,2	43,5
80	In my life	45,5	54,2	63,6	52,6	68,2	65,4	74,2	72,3	74,5
81	Wait	33,1	43,5	48,2	22,1	51,5	49,2	60,4	57,9	57,5
82	If i needed someone	44,6	40,2	50,9	46,4	64,8	49,1	75,2	65,4	72,0
83	Run for your life	27,6	46,8	44,5	29,3	87,5	52,7	87,8	87,0	89,6
84	Taxman	50,3	35,8	50,1	30,1	51,1	39,9	84,6	63,8	75,2
85	Eleanor rigby	25,8	40,3	40,2	48,3	73,4	52,8	76,2	77,3	71,8
86	I'm only sleeping	24,8	41,1	41,3	15,3	48,8	52,2	58,2	55,9	55,8
87	Love you to	15,1	16,8	9,1	45,9	18,4	40,1	16,0	16,1	10,9
88	Here, there and everywhere	35,6	50,5	52,6	49,9	47,4	58,3	66,6	67,3	65,0
89	Yellow submarine	20,0	12,2	11,2	28,2	22,5	27,2	25,0	29,9	29,5
90	She said she said	41,8	38,6	41,5	65,7	65,2	60,7	71,5	74,7	74,0
91	Good day sunshine	41,9	41,2	49,8	55,6	70,9	51,2	76,0	75,3	72,1
92	And your bird can sing	68,5	57,4	64,8	70,3	76,8	64,1	86,1	79,6	84,4
93	For no one	22,1	37,4	42,6	26,5	53,7	42,3	60,5	63,0	66,5
94	Doctor robert	44,4	37,3	50,2	44,3	39,1	52,0	73,6	46,6	75,3
95	I want to tell you	66,7	53,4	70,2	66,6	79,9	70,0	74,7	75,9	80,8
96	Got to get you into my life	42,4	39,0	51,1	46,2	43,2	48,8	57,3	53,4	65,6
97	Tomorrow never knows	54,7	27,8	59,4	5,8	27,4	35,0	76,2	41,4	70,4
98	Sgt.pepper's lonely hearts club band	66,6	50,6	61,5	39,3	59,6	54,3	75,3	72,8	75,4
99	With a little help from my fr	46,4	52,0	62,0	41,3	72,5	67,7	83,6	77,5	81,9
100	Lucy in the sky with diamonds	31,6	46,8	50,5	26,3	60,3	51,9	68,2	65,3	65,6

Angaben in Prozent

Tabelle F.3 – Ergebnisse Akkorderkennung – Teil 3

Thesen:

- Mathematisch-geometrische Tonräume stellen ein wichtiges Element für die Analyse und Synthese von musikalischen Audiosignalen dar.
- Tonale Symmetrie ist eine wichtige Grundeigenschaft abendländischer Musik. Durch Beachtung der tonalen Symmetrie können bestehende Tonräume vereinfacht dargestellt und viele musikalische Zusammenhänge leichter berechnet werden.
- Bei der Entwicklung von Tonraummodellen muss zwischen tonartbezogenen und tonartübergreifenden Modellen unterschieden werden.
- Der tonartbezogene Terzkreis *TR* sowie der tonartbezogene Quintenzirkel *FR* des Symmetriemodells sind gut geeignet, um viele Eigenschaften abendländischer Tonart zu modellieren – wie z.B. die Stabilität von Tonigkeiten in einem tonalen Kontext, Strebewirkungen von Tonigkeiten und Diskriminierung des Tongeschlechts.
- Die enharmonische Identität einer Tonigkeit repräsentiert wichtige, nicht vernachlässigende Wahrnehmungseigenschaften. Im *TR* werden enharmonisch verwechsellte Tonigkeiten musiktheoretisch korrekt modelliert.
- Mathematisch-geometrische Tonräume sind gut geeignet, um neue innovative musikalische Mensch-Maschineschnittstellen zu entwickeln.
- Ein Tonraumsynthesizer basiert 1.) auf einer geometrischen Anordnung von Tönen im Tonraum und 2.) auf einem Raumausschnitt, welcher die zu spielenden Töne selektiert. Die Klangverformung erfolgt entweder durch Manipulation der Tonpositionen im Tonraum oder durch Verformung oder Transformation des selektierten Raumausschnittes.
- Der Tonigkeits-Tonhöhen-Raum ist ein wichtiger Tonraum, der aus einer Tonigkeitsachse und einer Tonhöhenachse besteht. Wird der selektierte Raumausschnitt eines Tonraumsynthesizers entlang der Tonigkeitsachse verschoben, so werden automatisch günstige Akkordverbindungen erzeugt.
- Die Berechnung eines Summenvektors im kreisförmigen Tonraum ist der entscheidende Schritt zur Analyse von musikalischen Ereignissen. Der Summenvektor beinhaltet wesentliche semantische Parameter.
- Maschinelle Lernverfahren, die sowohl mit einer funktionellen als auch einer zeitlichen Relation arbeiten, sind den Verfahren überlegen, die nur mit einer funktionellen Relation arbeiten.

-
- Für die Akkorderkennung ist die Kombination von *Hidden Markov Models* mit einem KNN-Wahrscheinlichkeitsschätzer wesentlich besser geeignet als mit einem Gauß-Wahrscheinlichkeitsschätzer.
 - Das Symmetriemodell ordnet musikalische Ereignisse wie z.B. Töne, Intervalle und Akkorde bei einem gegebenen Kontext funktionsharmonisch richtig zu.
 - Bei der Tonartanalyse ist es sinnvoller in die sechs gebräuchlichen Kirchentonarten (Ionisch, Dorisch, Phrygisch, Lydisch, Mixolydisch und Aeolisch) zu unterscheiden anstatt lediglich in Dur und Moll.
 - Der tonartbezogene Quintenzirkel *FR* des Symmetriemodells trennt Dur- und Mollereignisse voneinander.
 - Der Ausgabevektor des Symmetriemodells ist für die Akkordanalyse besser geeignet als der Chromavektor, der dem Stand der Technik entspricht.