Merkmalsmodellierung für das maschinelle Lernen von Musikstilen

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

der Fakultät für Informatik

der Universität Fridericiana zu Karlsruhe (TH)

genehmigte

D i s s e r t a t i o n

von

Karin Höthker

aus Marburg

Tag der mündlichen Prüfung:	13. Juli 2005
Erster Gutachter:	Prof. Dr. Wolfram Menzel
Zweiter Gutachter:	Prof. Dr. Werner Zorn

Inhaltsverzeichnis

1 Einleitung **1**

1.1 Beispiel . . . 1

1.2 Ziele . . . 3

1.2.1 Informatische Ziele . . . 3

1.2.2 Musikwissenschaftliche Ziele . . . 4

1.3 Repräsentation zeitabhängiger Merkmale . . . 6

1.4 Stilanalyse . . . 9

1.5 Aufbau der Arbeit . . . 11

2 Grundlagen **13**

2.1 Informatisch-mathematische Grundlagen . . . 13

2.1.1 Reguläre Ausdrücke . . . 13

2.1.2 Grammatiken . . . 14

2.1.3 Graphen . . . 19

2.1.4 Neuronale Netze . . . 22

2.1.5 Merkmalsselektion . . . 27

2.2 Musikalische Grundlagen . . . 31

2.2.1 Melodie . . . 32

2.2.2 Musikalischer Stil . . . 33

2.2.3 Beispiele für musikalische Merkmale . . . 35

3 Stand der Forschung **39**

3.1 Andere Verfahren des maschinellen Lernens 40

3.1.1 Neuronale Netze und Support-Vektor-Maschinen 40

3.1.2 Verfahren zur automatischen Merkmalsextraktion 41

3.2 Repräsentation und Transformation musikalischer Strukturen 42

3.2.1 Datenformate zur Repräsentation von Musikdaten 42

3.2.2 Analytische Musikstrukturmodelle 43

3.3 Stilmodellierung 44

3.3.1 Musikalische Merkmalsselektion 44

3.3.2 Melodiegenerierung 45

3.3.3 Stilunterscheidung und Stilerkennung 46

3.4 Vorarbeiten der eigenen Forschungsgruppe 46

3.4.1 Methodischer Ansatz: Analyse durch Synthese 46

3.4.2 Analyse von Merkmalen 46

3.4.3 Lernen musikalischer Strukturen mit maschinellen Verfahren 47

3.4.4 Verfahren zur Melodiegenerierung 48

3.4.5 Evaluierung durch Stilerkennung 50

4 Repräsentation musikalischer Strukturen **53**

4.1 Musikalische Zeit 55

4.2 Musikalische Objekte 60

4.2.1 Beispiel zur Strukturierung von Problemwissen 60

4.2.2 Objektgrammatik 70

4.2.3 Typgrammatik 76

4.2.4 Templategrammatik 80

4.2.4.1 Typbedingungen 88

4.2.4.2 Praktische Umsetzung der Expansion 90

4.2.5 Erweiterbare Zeitreihengrammatik 93

4.2.6 Beispiel für eine Erweiterbare Zeitreihengrammatik 98

4.3 Musikalische Strukturen 105

5 Transformation musikalischer Strukturen 107

5.1 Operator 110

5.2 Operatoren für Zeitreihen 116

5.2.1 Navigation 116

5.2.1.1 Iterator 118

5.2.1.2 Beispiele 120

5.2.2 Gleichzeitigkeit 123

5.2.2.1 Koinzidenzoperator 123

5.2.2.2 Beispiele 124

5.2.3 Anwendungswissen 127

5.2.3.1 Wissensoperator 128

5.2.3.2 Beispiele für allgemeine Wissensoperatoren 128

5.2.3.3 Codierungs- und Decodierungsoperatoren 129

5.2.3.4 Neurooperatoren 136

5.2.4 Strukturmanipulation 143

5.2.4.1 Zerlegung von Objekten 144

5.2.4.2 Konstruktion von Objekten 145

5.2.4.3 Weitere Beispiele 147

5.3 Operatorgraph 153

5.4 Anwendungen von Operatorgraphen 159

5.4.1 Durchlaufstrategie 159

5.4.2 Berechnung musikalischer Merkmale 160

5.4.3 Berechnung von Lernmustern 164

5.4.4 Auswertung von Klassifikatoren 165

5.4.5 Evolutionäre Generierung von Melodien 167

5.5 Operatorgraph und Zeitreihengrammatik 169

6 Musikalische Stilmodellierung **171**

6.1 Merkmalsselektion 172

6.1.1 Modellierung eines einzelnen Stils 172

6.1.2 Stilunterscheidung 185

6.1.3 Stilerkennung 194

6.2 Evolutionäre Komposition 197

6.2.1 Vorgehensweise 198

6.2.2 Ein einfaches Beispiel 205

6.2.3 Ein komplexeres Beispiel 214

6.3 Merkmalssuche 218

A Das System MeloLab **225**

Kapitel 1

Einleitung

Gegenstand dieser Arbeit ist die Analyse von Melodien mit Methoden des maschinellen Lernens.

1.1 Beispiel

Eine Melodie lässt sich auf die einfachste Weise als eine Folge von Tönen beschreiben. Ein Beispiel zeigt jedoch, dass die Darstellung einer Melodie als Tonfolge die strukturellen Verflechtungen erfasst, die innerhalb einer Melodie auftreten können. Im Thema des *Largo* aus dem *Winter* der "Vier Jahreszeiten" von Antonio Vivaldi durchdringen einander verschiedene melodische und rhythmische Bewegungen (Abbildung 1.1). In Takt 3-4 sind zwei gegenläufige Bewegungen miteinander verschränkt: Aufsteigende melodische Linien bilden eine Sequenz, die sich auf einer übergeordneten Zeitebene abwärts bewegt. Ähnlich in Takt 5-6: Dort schwingen sich Wechselnotenmotive zu Zieltönen auf, die auf einer übergeordneten Zeitebene eine aufsteigende Linie bilden. Das auftaktige rhythmische Motiv 𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯 und seine Abwandlung 𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯 ziehen sich durch die ganze Melodie und rufen damit einen Eindruck rhythmischer Geschlossenheit hervor.

Diese Beobachtungen zeigen, dass die Beschreibung einer Melodie lediglich als Tonfolge – wie in bisherigen Ansätzen der lernbasierten Melodiemodellierung weitestgehend der Fall – dem

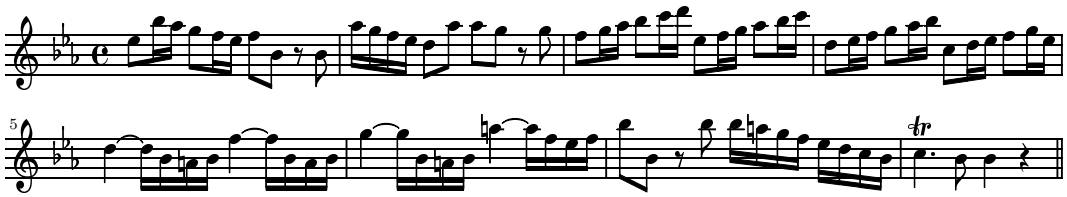

Abbildung 1.1: Antonio Vivaldi: Largo (Vier Jahreszeiten – Winter)

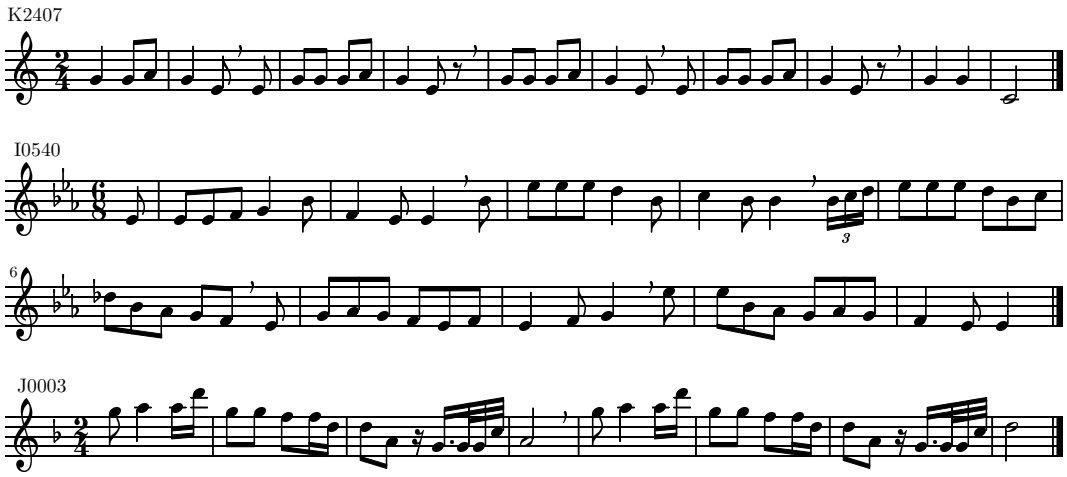

Abbildung 1.2: Drei Melodien aus der Essener Volksliedsammlung K2407: Kinderlied “Nix in der Grube”, I0540: Irisches Liebeslied, J0003: Volkslied aus der chinesischen Provinz Shanxi

komplexen Beziehungsgeflecht ihrer Teilstrukturen nicht gerecht wird. Bei einer computerunterstützten Analyse musikalischer Stile, die auf die Untersuchung der inneren musikalischen Zusammenhänge von Melodien abzielt, ist es daher wichtig, musikalische Struktureigenschaften (sogenannte *Merkmale*) explizit zu repräsentieren.

Es ist nicht möglich, eine Repräsentation zu entwerfen, die alle denkbaren melodischen Merkmale abdeckt, da diese zu zahlreich sind und abhängig von Fragestellung und untersuchtem Musikstil variieren. Das oben diskutierte Beispiel zeigt aber einige allgemeine Prinzipien auf, die beim Entwurf einer Repräsentation musikalischer Merkmale berücksichtigt werden sollten: Musikalische Ereignisse wie melodische Bewegung finden auf mehreren Zeitebenen statt. Musikalische Ereignisse innerhalb einer Zeitebene müssen nicht in gleichmäßigen Abständen angeordnet sein und können einander sogar überlappen. Charakteristische Strukturen wie z.B. ein rhythmisches Motiv durchziehen eine Melodie und setzen damit zeitlich entfernte Positionen einer Melodie zueinander in Beziehung. Die Analyse unterschiedlicher Merkmale (wie z.B. rhythmische Motive und Bewegungsrichtung) beleuchtet unterschiedliche Perspektiven und liefert so verschiedene Sichten auf dieselbe Melodie.

Ein zweites Beispiel illustriert die Bandbreite an möglichen Merkmalen bei der Untersuchung von Volksliedern unterschiedlicher Herkunft, die im Vergleich mit durchkomponierten Melodien wie z.B. dem barocken Thema in Abbildung 1.1 eine eher einfache Struktur besitzen. Abbildung 1.2 zeigt drei Volkslieder aus verschiedenen Melodiecorpora der Essener Volksliedsammlung [19], einer Sammlung von etwa 10000 Volksliedern europäischer und chinesischer Herkunft, die unter Leitung des Musikwissenschaftlers Helmut Schaffrath zusammengetragen wurde.

Die erste Melodie K2407 ist ein Kinderlied aus Thüringen und Schlesien, das bei Tanz- und Laufspielen gesungen wurde. Die leicht variierte Wiederholung der kurzen zweitaktigen

Phrasen, die ihrerseits Tonwiederholungen enthalten, die Konzentration der Melodie auf den Terz/Quintbereich der Skala und der bestätigende Quintfall in den letzten beiden Takten verleihen der Melodie den Charakter eines Abzählreims.

Die zweite Melodie I0540 ist ein irisches Liebeslied, dessen melodische Entwicklung großzügiger angelegt ist als beim Kinderlied K2047. Auch hier sind die Phrasen zweitaktig, sie erscheinen jedoch durch den metrisch zweigeteilten $\frac{6}{8}$-Takt länger als bei der ersten Melodie im $\frac{2}{4}$-Takt. Jede Phrase entwickelt sich aus ihrer Vorgängerin, indem sie rhythmische und melodische Elemente aufgreift (z.B. Dreifachrepetition des Tons *es* in den Takten 3 und 5, Sekundschritte mit Richtungswechsel in den Takten 7 und 9). Der triolische Auftakt zu Takt 5 wäre in einem Kinderlied untypisch, ebenso die tiefalterierte siebte Stufe *des* in Takt 6, die eine Harmonisierung des Takts mit der Harmoniefolge Doppelsubdominante-Tonika nahelegt.

Die dritte Melodie J0003 ist ein Volkslied aus der chinesischen Provinz Shanxi. Die Melodie besteht aus zwei viertaktigen Phrasen und gehört damit zu den kurzen Shanxi-Melodien der Sammlung. Die zweite Phrase wiederholt die erste, endet aber auf *d* statt auf *a*. Eine Verzierung führt zu den Schlusstönen beider Phrasen hin. Die Folge von schneller Quarte und Quinte in Gegenbewegung an einer metrisch unbetonten Position (Takt 1 und 5) hat ebenfalls Verzierungscharakter und würde einen durchschnittlichen europäischen Volksliedsänger überfordern. Der Tonvorrat der Melodie ist die pentatonische Skala *f-g-a-c-d*. Im Vergleich zu den beiden anderen Melodien springen der große Ambitus (Duodezime) und die hohe Lage (Spitzenton: d''') ins Auge.

Die drei diskutierten Melodien weisen also unterschiedliche Merkmale auf, die man bei der Charakterisierung der zugehörigen Volksliedstile als Kandidaten für stilrelevante Merkmale verwenden kann. Gestützt auf eine größere Zahl von Melodiebeispielen kann dann mit Hilfe lernbasierter Methoden die statistische Relevanz der Kandidaten untersucht werden.

1.2 Ziele

1.2.1 Informatische Ziele

Repräsentation zeitabhängiger Merkmale. Aus Sicht der Informatik handelt es sich bei den Musikbeispielen aus dem vorhergehenden Abschnitt um diskrete Zeitreihen mit einer komplexen inneren Struktur. Wenn diese innere Struktur nur implizit gegeben ist (wie z.B. bei der Repräsentation von Melodien als Tonfolgen), ist es wenig aussichtsreich, Methoden des maschinellen Lernens einzusetzen, da sie nicht über das problemspezifische Wissen verfügen, das nötig wäre, um relevante Strukturzusammenhänge aus den Basisdaten zu extrahieren. Man benötigt daher Vorverarbeitungsverfahren, mit denen sich die Daten so aufbereiten lassen, dass den maschinellen Lernverfahren durch die Ausgangsdaten implizierte Merkmale explizit zur Verfügung stehen.

Ein informatisches Ziel dieser Arbeit besteht darin, eine Repräsentation für zeitabhängige Merkmale zu entwickeln, die die Aufbereitung diskreter Zeitreihen für das maschinelle Lernen erlaubt. Dabei ist es wichtig, dass die Berechnung von Merkmalen und Lernmustern

automatisch erfolgt, da beim maschinellen Lernen zahlreiche Trainingsbeispiele benötigt werden, die nicht mit akzeptablem Aufwand manuell bearbeitet werden können.

Wenn man diskrete Zeitreihen unterschiedlicher Herkunft untersuchen will, muss das Problemwissen einer solchen Merkmalsrepräsentation austauschbar sein. Die Beispiele im ersten Abschnitt haben gezeigt, dass die Modellierung jedes musikalischen Stils andere Merkmale erfordert. Darüberhinaus soll der entwickelte Ansatz so allgemein sein, dass er auch auf nicht-musikalische diskrete Zeitreihen wie z.B. Finanzzeitreihen anwendbar ist.

Für Algorithmen wie das Trainieren von Klassifikatoren, die Merkmalsselektion oder die musikalischen Stilunterscheidung, die die aufbereiteten Zeitreihen als Eingabe nutzen, sollte das verwendete Problemwissen nicht sichtbar sein, damit sie unabhängig von einer konkreten Aufgabenstellung implementiert werden können. Merkmale werden dann als Variable der Algorithmen behandelt. Auf diese Weise entstehen Familien von Algorithmen, die gleich arbeiten, aber unterschiedliche Merkmale verwenden. Dies stellt eine Verallgemeinerung musikspezifischer Algorithmen dar, deren Merkmale üblicherweise fest implementiert sind.

Der Nutzen der hier entwickelten Repräsentation für das maschinelle Lernen liegt also darin, dass sie nicht nur für musikalische, sondern für beliebige Merkmale eine Aufbereitung diskreter Zeitreihen mit komplexer innerer Struktur für das maschinelle Lernen ermöglicht. Dabei werden insbesondere variable Zeitebenen berücksichtigt, die sich aus unterschiedlichen Interpretationen der Daten ergeben. Finanzzeitreihen und Absatzprognosen sind nicht-musikalische Beispiele, bei denen die Betrachtung variabler Zeitebenen sinnvoll ist. Ragg [70] sagt z.B. Absatzzahlen für BILD-Zeitungen vorher, indem er Differenzen von Absatzzahlen im Wochenabstand berechnet. Dabei müssen Feiertage und besondere Ereignisse berücksichtigt werden.

Melodien sind aus der informatischen Perspektive deshalb als Untersuchungsgegenstand interessant, weil sie bei beschränktem Ausgangsmaterial besonders komplexe innere Strukturen aufweisen, die mit den bisherigen Ansätzen nicht hinreichend modelliert werden können.

1.2.2 Musikwissenschaftliche Ziele

In Abschnitt 1.1 wurden ein barockes Thema und drei Volkslieder unterschiedlicher Herkunft diskutiert. Ob die dort identifizierten Eigenschaften auch für stilistisch ähnliche Melodien gelten, lässt sich aber nicht aus einer Einzelfallbetrachtung, sondern nur anhand einer statistisch abgesicherten Analyse einer großen Zahl von Beispielen des gleichen Stils schließen. In der vorliegenden Arbeit soll untersucht werden, wie sich Verfahren des maschinellen Lernens auf die musikalische Stilanalyse anwenden lassen.

Der Einsatz maschineller Lernverfahren bietet sich besonders bei Problemen an, die sich schlecht durch Regeln formalisieren lassen, für die aber viele Lösungsbeispiele existieren. In der Musik gibt es zum Beispiel kein allgemein akzeptiertes Regelwerk zur Melodiekomposition, während Stimmführungsregeln für den traditionellen vierstimmigen Chorsatz zum musikwissenschaftlichen Grundwissen gehören.

Auffinden stiltypischer Merkmale in Melodien. In dieser Arbeit werden drei Stilmodellierungaufgaben mit Methoden des maschinellen Lernens untersucht. Bei der *Einzelstilmodellierung* werden Merkmale bestimmt, die im Kontext eines vorgegebenen Musikstils

für die Vorhersage eines weiteren Merkmals relevant sind. Bei der *Stilunterscheidung* besteht die Aufgabe darin, Merkmale zu finden, die einem lernbasierten System die Zuordnung von Musikbeispielen zu einem von zwei vorgegebenen Stilen erlauben. Die *Stilerkennung* verallgemeinert die Stilunterscheidung so, dass eine beliebige Anzahl von Stilen vorgegeben werden kann.

Eine Gemeinsamkeit dieser drei Aufgabenstellungen besteht darin, dass zur Auswahl lösungsrelevanter Merkmale ein objektives Gütekriterium zur Verfügung steht. Die Leistungsfähigkeit eines Stilunterscheiders wird z.B. ausgedrückt, indem man seine Generalisierungsleistung mit Hilfe der Stilunterscheidungsgüte auf einer Testmenge schätzt, die nicht zum Einstellen der Modellparameter verwendet wurde. Die Eingabemerkmale des leistungfähigsten Stilunterscheiders werden als besonders relevant für die Unterscheidung der untersuchten Stile ausgewählt.

Diese Vorgehensweise ermöglicht eine Analyse elementarer Zusammenhänge zwischen musikalischen Merkmalen und die automatische Optimierung lernbasierter Modelle zur Prognose musikalischer Merkmale.

Generierung neuer Melodien. Die Prognose musikalischer Merkmale lässt sich auch zur Generierung neuer Melodien einsetzen. Es stellt sich jedoch die Frage, ob ein Modell, das Melodien mit einem Ensemble objektiver Bewerter generiert, tatsächlich das musikalisch-ästhetische Konzept annähert, das Musikexperten oder Versuchspersonen von einem Musikstil haben. Genau genommen lernen die bisher angesprochenen Modelle, die Werte eines musikalischen Merkmals zu prognostizieren. Damit erfasst man, ob das betreffende Merkmal aus den Beispielen für den betrachteten Stil überhaupt gelernt werden kann, und von welchen anderen Merkmalen es abhängt. Die Identifizierung solcher "lernbaren" Merkmale ist eine notwendige Voraussetzung für die Erstellung eines lernbasierten Stilmodells; hinreichend für die Charakterisierung eines Musikstils ist sie aber nicht.

Im Vergleich zu Problemstellungen mit objektiven Gütekriterien besteht die Besonderheit bei der Modellierung eines musikalischen Stils darin, dass das musikalisch-ästhetische Zielkonzept – der musikalische Stil – sich nicht in Regeln fassen lässt. Vielmehr wird angestrebt, kognitive Konzepte eines Musikstils zu approximieren, die Musikexperten durch Studium und Hörerfahrung entwickelt haben. Um ein lernbasiertes Stilmodell mit diesem Optimierungsziel evaluieren zu können, muss das enthaltene Wissen in einer Form aufbereitet werden, die sich mit dem kognitiven Stilkonzept von Musikexperten in Verbindung bringen lässt. Hier bietet sich die Generierung neuer Melodien mit Hilfe eines lernbasierten Modells an, denn die Stiltreue generierter Melodien kann von Musikexperten oder Versuchspersonen besser bewertet werden als eine Menge von Merkmalen, die von einem Algorithmus als relevant eingestuft wurden. Die Gesamterscheinung einer Melodie spiegelt wider, ob die lernbasierten Komponenten des Stilmodells die wesentlichen Aspekte des Stils erfassen.

Ein methodischer Vorteil bei dieser Vorgehensweise besteht darin, dass man keinerlei Annahmen über die Funktionsweise des kognitiven Stilmodells der Musikexperten machen muss. Demgegenüber wäre eine Bewertung von Mengen selektierter Merkmale durch Experten, wie sie die Einzelstilmodellierung, Stilunterscheidung und Stilerkennung liefern, methodisch fragwürdiger, da die selektierten Merkmale im Kontext des kognitiven Stilkonzepts eines Experten möglicherweise keine Bedeutung haben oder sogar das Stilkonzept des Experten beeinflussen.

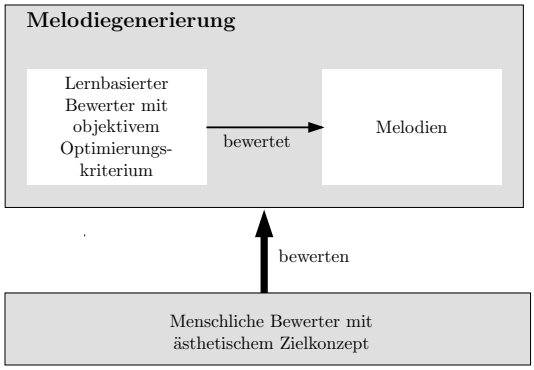

Abbildung 1.3: Zweistufige Bewertung von Melodien bei der Melodiegenerierung mit einem lernbasierten Modell.

Insgesamt wird ein lernbasiertes Stilmodell mit einem zweistufigen Ansatz bewertet (Abbildung 1.3): Es werden Melodien generiert, die ein lernbasierter Bewerter mit einem objektiven Gütekriterium optimiert. In einem übergeordneten Schritt vergleichen Musikexperten die bei der Melodiegenerierung als optimal eingestuften Melodien mit ihrer ästhetischen Zielvorstellung. Die Zusammensetzung des lernbasierten Bewerters und die Generierungsstrategie werden so lange verändert, bis das Stilkonzept der Musikexperten getroffen wird oder keine Verbesserung mehr eintritt.

Für die Musikwissenschaft stellt die Anwendung lernbasierten Verfahren auf musikalische Fragestellungen ein Werkzeug dar, mit dem die Relevanz musikalischer Merkmale statistisch fundiert untersucht werden kann. Die Auswahl der potentiell relevanten Merkmale stellt dabei bereits eine Hypothese über die Eigenschaften des untersuchten Stils dar. Da bei der Modellierung werden die musikalischen Merkmale dann formal definiert. Dadurch werden alle musikalischen Annahmen offengelegt und einer Diskussion zugänglich gemacht.

1.3 Repräsentation zeitabhängiger Merkmale

In dieser Arbeit wird eine Repräsentation für diskrete Zeitreihen entwickelt, die speziell auf die Anforderungen des maschinellen Lernens zugeschnitten ist. Die Repräsentation stellt die Grunddaten einer Zeitreihe und daraus abgeleitete Analysen dar. Sie bildet sowohl die Grundlage für die Berechnung dieser Analysen als auch für die Generierung von Lernmustern und kann an das problemspezifische Wissen unterschiedlicher Fragestellungen angepasst werden. Merkmale einer diskreten Zeitreihe werden getrennt von den Daten der Zeitreihe explizit repräsentiert, so dass sie als Variablen in Algorithmen verwendet werden können.

Ansichten. Zeitreihen bestehen aus Ereignissen, die einen Startzeitpunkt, eine Dauer und einen problemspezifischen Wert besitzen. Merkmale charakterisieren Eigenschaften einer Zeitreihe, indem sie Zusammenhänge zwischen Ereignissen aus einer festgelegten Perspektive beschreiben. In Anlehnung an die unterschiedlichen Datensichten in Softwaresystemen

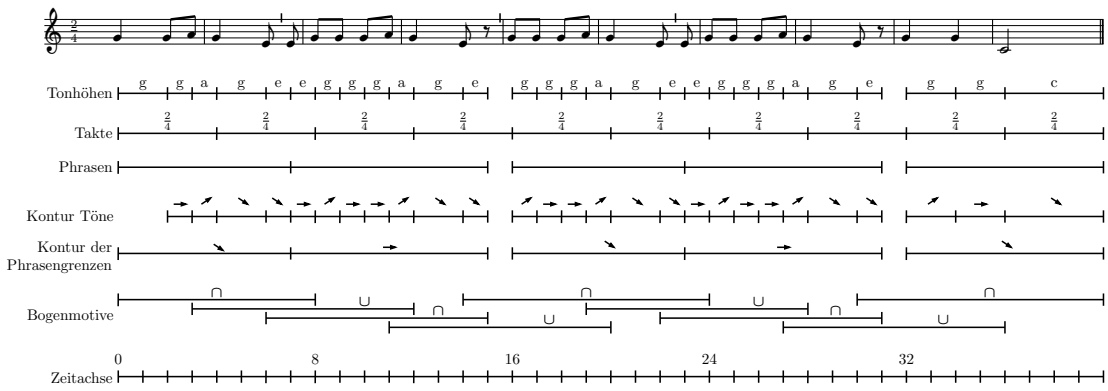

Abbildung 1.4: Fünf Melodieansichten des Kinderlieds “Nix in der Grube” (K2407)

wird die Darstellung einer Zeitreihe bezüglich eines Merkmals in dieser Arbeit als *Ansicht* (engl. *view*) der Zeitreihe bezeichnet.

Abbildung 1.4 zeigt ein Beispiel für ein Melodie, d.h. für eine musikalische Zeitreihe, und fünf Ansichten dieser Melodie. Eine Ansicht besteht aus musikalischen Ereignissen, die sich jeweils aus einem Zeitintervall (*Segment*) und einem musikalischen Symbol (*Wert*) zusammensetzen, wobei das Symbol auch fehlen kann. Die oberen drei Ansichten “Tonhöhen”, “Takte” und “Phrasen” geben die Tonhöhen, Takte und Phrasen der Melodie wieder. Sie ergeben sich direkt aus der Originaldarstellung der Melodie in der EsAC-Volkslied-Sammlung [19]. Die Phrasenansicht besteht aus Segmenten ohne musikalisches Symbol, da eine Phrase allein durch ihre zeitliche Position bestimmt wird. Die Taktansicht gibt die Position der einzelnen Takte wieder und könnte auch einen Auftakt oder Taktwechsel darstellen.

Die untere Ansicht “Bogenmotive” wurde aus der Tonhöhenansicht abgeleitet. Es handelt sich um eine einfache Analyse der Melodie, die die konvexen und konkaven Teilstücke der Melodie inklusive konstanter Randstücke wiedergibt. Diese Ansicht zeigt auch, dass bereits eine einfache Analyse einer Melodie musikalische Ereignisse beinhalten kann, die einander überlappen.

Jede Ansicht enthält musikalische Symbole eines Typs (z.B. Tonhöhen, Taktarten). Ein Beispiel zeigt, dass der Typ einer Ansicht noch nicht festlegt, welches Merkmal einer Melodie die Ansicht wiedergibt: In Abbildung 1.4 sind zwei Ansichten mit Kontursymbolen angegeben. Die erste Ansicht “Kontur Töne” zeigt die Intervallrichtung für aufeinanderfolgende Töne der Melodie. Die zweite Ansicht “Kontur der Phrasengrenzen” bildet die Kontur für den Anfangs- und Endton jeder Phrase und ordnet diese dem Segment der Phrase zu. Beide Ansichten spiegeln unterschiedliche Merkmale der Melodie wider.

Erweiterbare Zeitreihengrammatik. Bei der Beschreibung der Ansichten wurden verschiedene Typen musikalischer Symbole genannt (Tonhöhe, Kontur, Bogenmotiv, Taktart). Die Wahl dieser Typen hängt von der untersuchten Fragestellung ab. Die Modellierung der zeitlichen Zuordnung der Elemente einer Zeitreihe ist hingegen bei allen Zeitreihenproblemen gleich. Als Grundlage für die Definition von Ansichten und die Repräsentation von

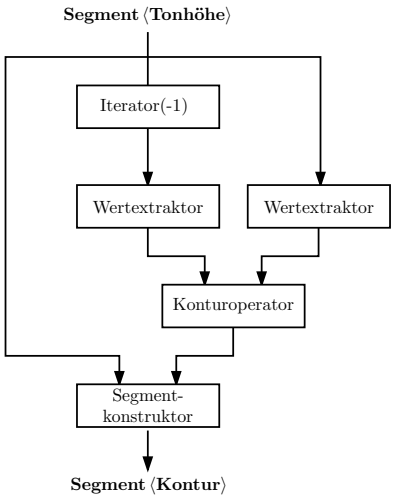

Abbildung 1.5: Operatorgraph zur Berechnung der Ansicht "Kontur Töne" aus Abb. 1.4

Merkmalen wird in dieser Arbeit eine *erweiterbare Zeitreihengrammatik* entwickelt, mit der sich die Symboltypen spezifizieren und generieren, auf denen die weiteren Modelle dieser Arbeit aufbauen. Die Zeitreihengrammatik ist erweiterbar, weil sie einen zeitreihenspezifischen Anteil vorgibt und durch die Definition eines problemspezifischen Anteils an das Anwendungsgebiet angepasst wird.

Operatorgraphen. Die beiden Ansichten vom Typ "Kontur" in Abbildung 1.4 zeigen, dass der Typ der Ansicht noch nicht festlegt, welches Merkmal einer Zeitreihe eine Ansicht wiedergibt. Vielmehr können sich selbst gleiche Ansichten in ihrer Bedeutung unterscheiden, wenn sie auf die Verknüpfung unterschiedlicher Information zurückgehen.

Um diese Mehrdeutigkeit von Ansichten zu berücksichtigen, werden Merkmale in dieser Arbeit getrennt von der Repräsentation der Zeitreihen definiert. Mit *Operatorgraphen* wird ein neues Beschreibungsmittel für Merkmale eingeführt, mit dem Berechnungsvorschriften für Symbole definiert werden. Definieren zwei Operatorgraphen unterschiedliche Funktionen, so haben auch von ihnen erzeugten Ansichten eine unterschiedliche Bedeutung, selbst wenn sie in einem konkreten Beispiel die gleichen Informationen enthalten.

Abbildung 1.5 zeigt einen Operatorgraphen, mit dem man die Kontur benachbarter Tonhöhen in Abbildung 1.4 berechnen kann. Der Operatorgraph setzt sich aus atomaren Aktionen zusammen, den sogenannten *Operatoren*, die zu einem zyklenfreien Graphen verkettet werden. Ein Segment einer Ansicht dient als Eingabe des Operatorgraphen und wird in der durch die Pfeile angegebenen Reihenfolge durch die Operatoren abgebildet und an die nachfolgenden Operatoren weitergereicht.

Im Beispiel werden die Segmente der Tonhöhenansicht aus Abbildung 1.4 als Eingabe des Operatorgraphen verwendet. Von der Eingabe ausgehend sucht der Iterator den linken Nachbarn des Eingabesegments auf. Aus diesem und aus der Eingabe werden mit Wertextrak-

toren die musikalischen Symbole extrahiert (hier Tonhöhen) und an den Konturoperator weitergereicht. Der Segmentkonstruktor verknüpft das berechnete Kontursymbol mit der Zeitinformation des Eingabesegments. Die Ausgabe des Operatorgraphen wird in der Konturansicht gespeichert. Für den zweiten Ton der Melodie, eine Achtel mit Tonhöhe g, erhält man beispielsweise die Kontur gleichbleibend ($\rightarrow$).

Auch bei der Entwicklung der Operatorgraphen werden zeitreihenspezifische von problemspezifischen Aspekten getrennt, um die Integration von neuem Problemwissen in den Ansatz zu erleichtern. Im Beispiel gibt es einen problembezogenen Operator (den Konturoperator) und drei verschiedene zeitbezogenene Operatoren, die vom betrachteten Problemfeld unabhängig sind (Iterator, Wertextraktor und Segmentkonstruktor).

Die Berechnung von Ansichten ist nicht die einzige Einsatzmöglichkeit für Operatorgraphen. So können auch Lernmuster mit Hilfe von Operatorgraphen generiert werden. Der Operatorgraph liefert dann eine numerische Ausgabe, die nicht in einer Ansicht der Datenrepräsentation, sondern in einer Lernmusterdatei abgelegt wird.

Besonders im Vergleich zur festen Implementierung musikalischer Transformationen bietet die Verwendung von Operatorgraphen den methodisch wichtigen Vorteil, dass Operatorgraphen unabhängig von ihrer Struktur immer auf dieselbe Weise ausgewertet werden. Es ist daher möglich, Algorithmen mit frei austauschbaren, von Operatorgraphen dargestellten Merkmalen zu implementieren, ohne dass auf die tatsächliche Beschaffenheit der Merkmale Bezug genommen werden muss. Operatorgraphen bilden damit eine einheitliche Schnittstelle zwischen einem generischen Algorithmus einerseits und einer problembezogenen Repräsentation musikalischer Daten andererseits.

1.4 Stilanalyse

Verallgemeinerte Musikrepräsentation. Während in früheren lernbasierten Ansätzen die Anzahl der in Frage kommenden Merkmale mit vereinfachenden Annahmen reduziert wurde, ist durch die variable Zeitdarstellung und die Austauschbarkeit von Problemwissen die Untersuchung komplexer musikalischer Strukturen nun besser möglich. Beispielsweise beziehen die neuronalen Netze bei der Choralharmonisierung in [32] ihre Eingabemerkmale nur aus einem lokalen Zeitfenster. Bei der Melodieergänzung von Kinderliedern in [89] motivierte der regelmäßige Aufbau der Melodien eine Zerlegung in Zeitsegmente gleicher Länge, wodurch die Menge der darstellbaren Merkmale eingeschränkt wurde. Diese Einschränkungen sind mit dem hier entwickelten Ansatz nicht notwendig.

Merkmalsselektion. Die mit der Verallgemeinerung der Repräsentation einhergehende deutliche Zunahme darstellbarer Merkmale macht es erforderlich, statt der bisherigen manuellen Merkmalsauswahl Merkmalsselektionsverfahren einzusetzen. Ein Merkmalsselektionsverfahren wählt aus einer vorgegebenen Menge von Merkmalen diejenigen aus, die für die Lösung einer Lernaufgabe relevant sind. Die gewonnenen Merkmale können aus Ausgangsmaterial für die Melodiegenerierung genutzt werden.

Melodiegenerierung. Wenn man musikalische Merkmale finden will, die einen musikalischen Stil angemessen beschreiben, geht man von der Grundannahme aus, dass maschinell

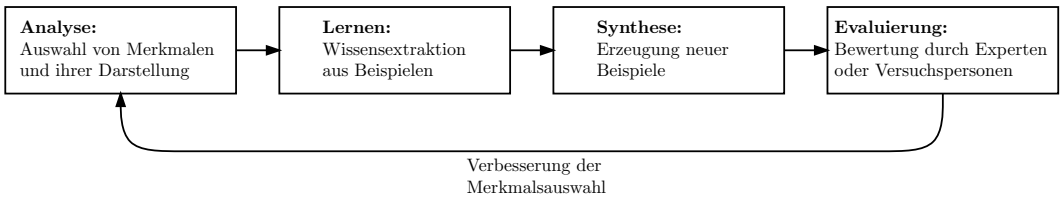

Abbildung 1.6: Zyklische Modellentwicklung mit dem Ansatz "Analyse durch Synthese"

lernbare Zusammenhänge auf die Relevanz der verwendeten Merkmale hindeuten. Der Ansatz der "Analyse durch Synthese" [101] nutzt diese Annahme zur zyklischen Verbesserung eines musikalischen Modells durch die Generierung neuer Musikstücke (Abbildung 1.6). Dabei ist ein lernbasiertes Modell mit austauschbaren Merkmalen gegeben. Nachdem man Merkmale ausgewählt und die Trainingsbeispiele entsprechend vorverarbeitet hat (*Analyse*), werden die lernbasierten Komponenten des Modells trainiert (*Lernen*). Bei der *Synthese* generiert man neue Musikstücke, deren Stiltreue bei der *Evaluierung* durch Experten oder Versuchspersonen bewertet wird. Durch die iterative Veränderung der verwendeten Merkmale und die Variierung der lernbasierten Bewerter wird die Qualität der erzeugten Stücke nach und nach verbessert. Die Generierung von Melodien dient also im Kontext dieser Arbeit der Stilanalyse, nicht der Komposition als Selbstzweck.

Operatorgraphen. Bei den Verfahren zur Stilanalyse, die in dieser Arbeit entwickelt werden, erfüllen die Operatorgraphen unterschiedliche Aufgaben. In allen Verfahren, die in Abschnitt 1.2.2 skizziert wurden, werden Operatorgraphen zur Vorverarbeitung der Lernbeispiele und zur Erzeugung von Lernmustern eingesetzt. Während die drei auf Merkmalsselektion aufgebauten Verfahren Einzelstilmodellierung, Stilunterscheidung und Stilerkennung bei der Optimierung nur auf die im Vorfeld generierten Muster zugreifen, spielen die Operatorgraphen bei der evolutionären Melodiegenerierung eine bedeutendere Rolle. Hier werden sie während der evolutionären Optimierung eingesetzt, um zufällige Melodien zu erzeugen, Melodien durch Mutation zu verändern und sie zu bewerten. Die Bewertung von Melodien mit Operatorgraphen setzt *Neurooperatoren* ein, also Operatoren, die ein neuronales Netz auswerten. Operatorgraphen erweisen sich damit als ein Werkzeug, mit dem die Interaktion lernbasierter Klassifikatoren untereinander und mit zusätzlichem Problemwissen spezifiziert werden kann. Auch eine hybride Bewertung von Melodien, die Regelwissen und lernbasierte Elemente verknüpft, wird auf diese Weise ermöglicht.

Das System MeloLab**.** Im Rahmen dieser Arbeit wurde das System MeloLab, eine flexible Entwicklungsumgebung zur Analyse und Synthese melodischer Strukturen, realisiert. Dort wurden die hier entwickelte Zeitreihenrepräsentation mit Ansichten und Operatorgraphen, sowie die untersuchten Stilanalyse- und Melodiegenerierungsverfahren implementiert.

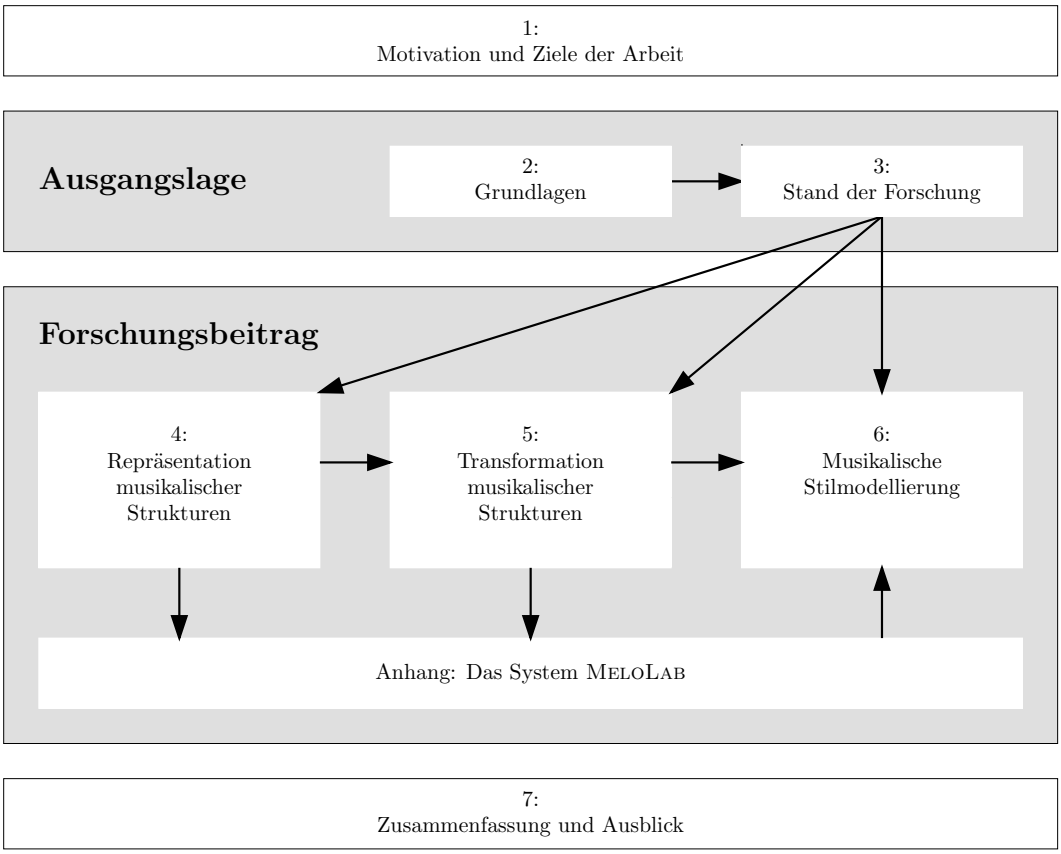

Abbildung 1.7: Aufbau der Arbeit

1.5 Aufbau der Arbeit

Der Rest der Arbeit gliedert sich wie in Abbildung 1.7 skizziert. In Kapitel 2 werden die informatischen und musikalischen Grundlagen vermittelt, die zum weiteren Verständnis notwendig sind. Das Kapitel ist als Referenz angelegt und kann beim Lesen zunächst übersprungen werden. Kapitel 3 ordnet die Arbeit in den gegenwärtigen Stand der Forschung ein. In Kapitel 4 wird eine allgemeine Repräsentation für musikalische Strukturen entwickelt. Auf dieser Basis werden in Kapitel 5 Operatorgraphen eingeführt, mit denen sich diskrete Zeitreihen unter Berücksichtigung von Hintergrundwissen für das maschinelle Lernen aufbereiten lassen. In Kapitel 6 wird gezeigt, wie Operatorgraphen bei der Merkmalsselektion und der Stilerkennung eingesetzt werden können. Die beschriebenen Algorithmen werden experimentell evaluiert. Die Ergebnisse werden in Kapitel 6.3 zusammengefasst. Anhang A skizziert das explorative neuronale System MeloLab, das den neuen Ansatz implementiert.