

**Variabilität und Konstanz von Melodien in
der Erinnerung. Ein Beitrag zur
musikpsychologischen Gedächtnisforschung**

**Dissertation
zur Erlangung der Würde des
Doktors der Philosophie
der Universität Hamburg**

vorgelegt von
Daniel Th. Müllensiefen
aus Freiburg i.Br.

Hamburg 2004

Inhalt

1 Einleitung.....	1
1.1 Gegenstand und Anspruch dieser Arbeit	1
1.2 Behandelte und nicht behandelte Themen dieser Arbeit	7
1.3 Weitere Schritte auf diesem Forschungsansatz.....	11
2 Theorien des musikalischen und melodischen Gedächtnisses.....	12
2.1 Modelle des musikalischen Arbeitsgedächtnisses	14
2.1.1 Das Modell des Arbeitsgedächtnisses nach Baddeley für melodische Inhalte	14
2.1.2 Das Modell des musikalischen Arbeitsgedächtnisses von William Berz	17
2.2 Übergreifende Modelle des musikalischen Gedächtnisses	20
2.2.1 Das Drei-Zeiten-Modell von Bob Snyder.....	20
3 Das Gedächtnis für Töne, kurze Tonfolgen und Melodien (Literaturübersicht)	27
3.1 Gedächtnis für Einzeltöne.....	27
3.1.1 Absolutes Gehör.....	27
3.1.2 Gedächtnis für Einzeltöne.....	34
3.1.3 Langzeitgedächtnis für absolute Tonhöhen echter Musik	41
3.1.4 Empirische Ergebnisse der Forschungen zum absoluten Gehör und zum Gedächtnis für Einzeltöne und absolute Tonhöhen ganzer Musikstücke (Tabelle)	46
3.2 Gedächtnis für kurze Tonfolgen	49
3.2.1 Bedeutung und untersuchungstheoretischer Ansatz	49
3.2.2 Eigenschaften von kurzen Tonfolgen als experimentelle Stimuli	50
3.2.3 Experimentelle und erkenntnistheoretische Grenzen.....	54
3.2.4 Untersuchte Variablen und experimentelle Paradigmen.....	57
3.2.4.1 Experimentelles Paradigma	57
3.2.4.2 Abhängige Variablen	59
3.2.4.2.1 Exkurs: Die Maße d' und AUC	60
3.2.4.2.2 Kritik der für die abhängigen Variablen verwendeten Maße.....	65
3.2.4.3 Unabhängige Variablen	66
3.2.5 Zusammenfassung der methodischen Schwierigkeiten der Experimente mit kurzen Tonfolgen.....	70
3.2.6 Ergebnisse der experimentellen Literatur mit kurzen Tonfolgen	72
3.2.6.1 Melodische Kontur.....	72
3.2.6.2 Melodische Struktur und regelgeleiteter Aufbau.....	77
3.2.6.3 Zusammenfassung der Erkenntnisse zum Gedächtnis für kurze Tonfolgen.....	82
3.2.7 Empirische Ergebnisse der Forschungen zum Gedächtnis für kurze Tonfolgen (Tabelle).....	85
3.3 Gedächtnis für Melodien.....	88
3.3.1 Wiedergabe unbekannter Melodien	89
3.3.2 Dimensionen der Speicherung von bekannten und unbekannten Melodien.....	95
3.3.2.1 Manipulation von Dimensionen bekannter Melodien.....	97
3.3.2.2 Speicherung von unbekanntem melodischen Material	100
3.3.3 Bedeutung von Akzenten und melodischer Erwartung für das Melodiegedächtnis ..	107
3.3.3.1 Das melodisch-rhythmische Konsonanzmodell von Caroline B. Monahan	108
3.3.3.2 Die Joint Accent Hypothesis von Boltz und Jones	116
3.3.3.3 Das Konzept der musikalischen Erwartungen von Mark A. Schmuckler	128
3.3.4 Melodieexperimente zur Überprüfung von allgemeineren Gedächtnistheorien.....	138
3.3.4.1 Dissoziation zwischen semantischem und episodischem Gedächtnis für Melodien	138
3.3.4.2 Untersuchungen zum Gedächtnis im Alter anhand von Melodien	140
3.3.5 Empirische Ergebnisse der Forschungen zum Gedächtnis für Melodien (Tabelle) ..	142

4 Melodische Ähnlichkeit	146
4.1 Melodische Ähnlichkeit als Werkzeug zur Analyse von Gedächtnisleistungen und zur Modellierung von Gedächtnismechanismen	146
4.1.1 Anwendungsfelder melodischer Ähnlichkeitsmodelle	146
4.1.2 Definition melodischer Ähnlichkeit als Grundlage für die Modelle des Melodiegedächtnisses	152
4.2 Die zu testenden Ähnlichkeitsmodelle	156
4.2.1 Vorstellung der Modelle	156
4.2.2 Zusammenfassung Ähnlichkeitsmaße	188
4.3 Erhebung subjektiver Ähnlichkeitsurteile	190
4.3.1 Das Hauptexperiment	190
4.3.1.1 Hypothesen zum Zusammenhang der menschlichen Ähnlichkeitsurteile mit den experimentellen Variablen	190
4.3.1.2 Methode	192
4.3.1.3 Ergebnisse	197
4.3.1.3.1 Übereinstimmung der menschlichen Ähnlichkeitsurteile und musikalische Expertise der selektierten Versuchspersonen	197
4.3.1.3.2 Zusammenhang zwischen musikalischer Expertise und der Erfüllung der Testkriterien	200
4.3.1.3.3 Abhängigkeit der Ähnlichkeitsurteile von den verwendeten Fehlertypen	208
4.3.1.3.4 Fehlerposition:	212
4.3.1.3.6 Einfluß der unterschiedlichen Melodien	219
4.3.2 Experiment 2 (Das Kontrollexperiment 1)	220
4.3.2.1 Hypothesen	221
4.3.2.2 Methode	222
4.3.2.3 Ergebnisse	224
4.3.2.3.1 Übereinstimmung in den Urteilen der Musikexperten	224
4.3.2.3.2 Beurteilungsunterschiede zwischen ‚richtigen‘ Varianten und ‚falschen‘ Varianten anderer Melodien	226
4.3.2.3.3 Zusammenhang der Ähnlichkeitsbewertungen in Experiment 1 und im Kontrollexperiment	227
4.3.3 Experiment 3 (Das Kontrollexperiment 2)	229
4.3.3.1 Methode	230
4.3.3.2 Ergebnisse	231
4.4 Modellierung menschlicher Urteile mit Ähnlichkeitsalgorithmen	232
4.4.1 Ermittlung des optimalen Ähnlichkeitsmaßes für das Hauptexperiment (Lineare Regression)	232
4.4.2 Ermittlung der optimalen Ähnlichkeitsmaße für das 1. Kontrollexperiment (Exp. 2)	238
4.4.2.1 Die ‚richtigen‘ Varianten:	238
4.4.2.2 Die ‚richtigen‘ und ‚falschen‘ Varianten:	241
4.4.3 Ermittlung des optimalen Ähnlichkeitsmaßes für das zweite Kontrollexperiment (Experiment 3)	245
4.4.4 Allgemeine Beobachtungen zum Verhalten und der Leistungsfähigkeit der getesteten Ähnlichkeitsmaße	247
4.5 Vergleich der Ähnlichkeitsmaße in Multidimensionaler Skalierung	250
4.5.1 Ansatz und Methode	250
4.5.2 Durchführung der MDS	254
4.6 Zusammenfassung	259
5 Wiedererkennen unbekannter Melodien	263
5.1 Vorüberlegungen und Modellierungsrahmen	263

5.2 Ökologische Validität von Design und Versuchsmaterialien	266
5.3 Methode	268
5.3.1 Versuchsteilnehmer.....	268
5.3.2 Materialien:	269
5.3.2.1 Die Teststücke:.....	269
5.3.2.2 Die Testmelodien:.....	273
5.3.3 Prozedur	277
5.3.4 Design und Hauptfaktoren	278
5.3.5 Weitere Einflussgrößen und Operationalisierungen	280
5.4 Datenauswertung und erste Ergebnisse	284
5.4.1 Richtige Identifizierung einzelner Teststücke	284
5.4.2 Subjektive Bekanntheit des Testmaterials	284
5.4.3 Verwendete Gedächtnisstrategien.....	286
5.4.4 Soziodemographische Beschreibung der Stichprobe.....	288
5.4.5 Die musikalische Vorbildung der Versuchsteilnehmer	290
5.4.5.1 Indexbildung und Datenverdichtung anhand der Variablen zur musikalischen Vorbildung	291
5.4.6 Gedächtnisleistungen der Versuchsteilnehmer	293
5.4.7 Unterschiede zwischen den Melodien	295
5.5 Statistische Modellierung.....	296
5.5.1 Datenaggregation mit Maßen aus der <i>Signal Detection Theory</i>	297
5.5.2 Vierfaktorielles Modell ohne Datenaggregation.....	303
5.6 Statistische Modellierung der <i>Versuchspersonenratings</i> (<i>data-mining</i> -Ansätze)	309
5.6.1 Variablenkodierung und Operationalisierung.....	310
5.6.1.1 Exkurs: Konstruktion des Prädiktors <i>Akzentähnlichkeit</i>	311
5.6.1.2 Die Prädiktorenliste	327
5.6.2 Das lineare Modell.....	328
5.6.2.1 Variablenselektion	328
5.6.2.2 Parameterschätzung	333
5.6.3 Die Baummodelle	337
5.6.3.1 Methode der Baummodelle.....	337
5.6.3.2 Der Klassifizierungsbaum.....	342
5.6.3.3 Der Regressionsbaum	345
5.6.3.4 Das zweistufige Klassifizierungsmodell.....	347
5.6.4 Evaluation der Modelle.....	351
5.6.4.1 Ergebnisse des linearen Modells.....	352
5.6.4.2 Ergebnisse des Klassifizierungsbaumes	353
5.6.4.3 Ergebnisse des Regressionsbaumes	353
5.6.4.4 Ergebnisse des zweistufigen Klassifizierungsbaumes	354
5.6.5 Diskussion der statistischen Modellierung	354
5.7 Zusammenfassung und Diskussion.....	364
6 Fazit.....	371
7 Literatur:	377
8 Anhang.....	397
A-4.3.1.2: Die 14 verwendeten Melodien der Experimente 1-3	397
A-4.3.1.2: Fragebogen zu den Experimenten 1-3 (nur Angaben zur Person, d.h. ohne Ratingskalen)	398
A-4.3.1.3.4 Modellierung der Ähnlichkeitsurteile der Experten anhand der Variablen des Versuchsdesigns.....	399
A-4.4.2 Kontrollexperiment.....	403
A-4.4.2.1 Die richtigen Varianten.....	403

A-4.4.2.2 Richtige und falsche Varianten.....	404
A-4.4.2.2 Plots zur Diskriminationsleistung der Ähnlichkeitsmaße.....	404
A-4.4.3 Experiment 3.....	406
A-5.4.7: Kreuztabelle zu Melodienummer und richtigen Antworten.....	408
A-5.5.2 Vierfaktorielles Modell ohne Datenaggregation: Geschätzte Randmittel und Scheffé-Tests.....	409
A-5.6.2.1 Vollständiges lineares Modell als Input zur Variablenselektion.....	413
A-5.6.3.2 Klassifizierungsbaum.....	414
A-5.6.3.3 Regressionsbaum.....	423
A-5.6.3.4 2-stufiger Klassifizierungsbaum:.....	433
A-5.6.4 Diskussion der statistischen Modellierung.....	439

1 Einleitung

1.1 Gegenstand und Anspruch dieser Arbeit

Das Problem ist alt, sehr alt. Schon Guido von Arezzo beklagt in seinem *Prologus in Antiphonarium* um 1025:

Temporibus nostris. super omnes homines (:)
fatui sunt cantores!
In omni enim arte. valde plura sunt,
quae nostro sensu cognoscimus:
quam ea quae a magistro didicimus.
Perlecto enim. solo psalterio (:)
omnium librorum lectiones. cognoscunt pueruli.

Et agriculturae scientiam: subito intelligunt rustici.
Qui enim (.)
unam vinam putare
unam arbusculam inserere,
unum asinum onerare. cognoverit:
- sicut in uno facit (:) - in omnibus similiter,
- aut etiam melius facere (:) -
- non dubitabit.

Miserabiles autem cantores. cantorumque discipuli!...
Etiamsi per centum annos. cotidie cantent,
numquam per se – sine magistro – unam
vel saltem parvulam – cantabunt antiphonam!
Tantum tempus. in cantando perdentes, in quanto /
et divinam, et saecularem scripturam /
potuissent plene cognoscere.¹

¹ Guido von Arezzo, *Prologus In Antiphonarium* (DMA.A.III, Edition von Joseph Smits van Waesberghe, S. 59-81; eigene Übersetzung):

In unserer Zeit sind von allen Menschen
die Sänger am bemitleidenswertesten.
In der ganzen Kunst gibt es nämlich viele Dinge,
die wir über unsere Sinne erkennen,
wie die, die wir vom Lehrer lernen.
Nachdem nämlich ein einziger Psalter gelernt ist,
lernen die Knaben die Lektionen aller Bücher kennen. //
Und die Wissenschaft vom Ackerbau verstehen die Landleute sofort.
Wer nämlich gelernt haben wird,
einen Weinstock zu beschneiden
ein Bäumchen einzupflanzen,
einen Esel zu beladen,
der wird nicht zögern,
so wie er es bei einem Mal macht,
es bei allen Gelegenheiten ähnlich zu tun,
oder sogar besser zu machen. //
Die Unglücklichsten sind aber die Sänger und die Gesangsschüler!...
Auch wenn sie 100 Jahre lang täglich sängen,
würden sie niemals aus sich heraus – ohne Lehrer – eine,
wenigstens eine kleine Antiphona singen!
Soviel Zeit, wie sie singend verlieren,
könnten sie sowohl die heilige, wie die weltliche Schrift
ganz erlernen.

Wir wissen leider nichts Genaues über Guidos Schüler, können aber auf Grund seiner langjährigen praktischen Erfahrung als Gesangs- und Musiklehrer annehmen, dass die Stichprobe, an der Guido seine empirischen Erkenntnisse zur Begrenztheit des menschlichen Melodiegedächtnisses gewann, nicht unbedingt klein war und hinsichtlich der zentralen Tendenz und der Streuung eine gute Schätzung der Gedächtnisfähigkeiten der Grundgesamtheit der männlichen gesangswilligen Bevölkerung seiner Zeit darstellt.

Demnach scheint die Fähigkeit des Menschen, zumindest in der europäischen Kultur, sich Melodien zu merken, also begrenzt zu sein. Zumindest war sie es vor 1000 Jahren, wenn man Guidos Klagen glauben darf. Guido, als einer der einflussreichsten Musikpädagogen seiner Zeit (Palisca, 1980), zog aus dieser allgemeinen menschlichen Gedächtnisschwäche für Melodien mehrere Konsequenzen.

Er propagierte die neue grafische Notenschrift auf Linien im Terzabstand als Gedächtnisstütze für bereits gelernte Gesänge und als Weg, neue Melodien vom Blatt zu singen, um so beim Einstudieren neuen Repertoires viel Zeit zu sparen. Zusätzlich entwickelte er das bis zum heutigen Tage gebrauchte Solmisationssystem, welches er in seiner *Epistula de ignoto cantu* beschreibt. Auch dieses System war als mnemonische Stütze für die Erinnerung von melodischen Strukturen gedacht. Es sollte Sängern helfen, die Lage des Halbtons und der Ganztonschritte in den seinerzeit gebräuchlichen sechsstufigen Tonleitern zu behalten. Nicht zuletzt wird Guido von Arezzo die sogenannte Guidonische Hand zugeschrieben, eine Visualisierung der Tonabstände in Tetrachorden, bei der den einzelnen Fingergliedern der linken Hand verschiedene Solmisationssilben zugeordnet sind.

Mit diesen Hilfssystemen trug Guido wesentlich zur Entwicklung der westlichen Vermittlungs- und Notationssysteme bei, mit denen heutzutage Musik unterrichtet wird und durch die Musiker in die Lage versetzt werden, auch sehr lange und komplexe Musikstücke vorzutragen zu können. Den Musikern, die Guido im Blick hatte, konnte also offensichtlich mit den in den letzten 1000 Jahren entwickelten Erinnerungshilfen und Gedächtnisstrategien bei ihrer täglichen Arbeit geholfen werden.

Wie ist es aber um das Gedächtnis für Melodien von Personen ohne spezielles musikalisches Training bestellt? Auch Konstrukteure von aktueller Musiksoftware stellen fest, dass das menschliche Gedächtnis für Melodien alles andere als perfekt ist und dies bei der Konstruktion von Schnittstellen zwischen Mensch und Musiksoftware besonders berücksichtigt werden muss (Pauws, 2002, S. 188):

It is obvious that people have imperfect memories for melodies [...]. This knowledge is indispensable for the development of 'query by humming' algorithms. [...]

Human memory is imperfect to recall all pitches in the right order. In our study, people sang just the line they remembered. They also added all kinds of ornaments [...] to beautify their singing or to ease muscular motor processes involved in singing.

Die Zitate von Stephen Pauws vom Philips Forschungslabor in Eindhoven und von Guido von Arezzo machen das Thema dieser Arbeit sehr anschaulich. Es geht um ein Problem, das schon vor langer Zeit in unserer Musikkultur zum Thema wurde und bis heute nur zum Teil als erforscht gelten kann: Melodien können im menschlichen Gedächtnis gespeichert werden, werden dabei aber verändert und teilweise vergessen. Für den Musikpsychologen stellt sich die Frage, ob der Transformations- und Vergessensprozess, den Melodien im Gedächtnis durchlaufen, ausschließlich dem Zufall folgt oder für jede Person vollständig individuell verläuft; oder ob sich – zumindest in einem gewissen Bereich – Zusammenhänge zwischen den Eigenschaften der Melodien, der Erfahrung der Personen, den Umständen der Speicher- und Erinnerungssituation und den Gedächtnisleistungen finden lassen.

Da diese Arbeit dem Bereich der kognitiven Musikpsychologie zuzuordnen ist und den dort vorherrschenden empirischen Ansatz verfolgt, sollen die im letzten Satz genannten Größen messbar und die Zusammenhänge zwischen ihnen quantifizierbar gemacht werden. Dies ist also Ziel und Anspruch dieser Arbeit: Es soll eine Beschreibung des menschlichen Gedächtnisses für Melodien gegeben werden, die aus quantifizierbaren Zusammenhängen zwischen messbaren Größen besteht.

Wechselt man vom Alltagsbegriff *Gedächtnis* zum wissenschaftlichen Terminus *Gedächtnis*, so wird eine allgemeingültige Definition oder Beschreibung desselben auf einmal sehr schwierig. Trotz der fast endlosen wissenschaftlichen Literatur zum Thema Gedächtnis bereitet die Ontologie des Gedächtnisses eine ganze Reihe von komplexen Problemen. Das 2001 von Nicolas Pethes und Jens Ruchatz herausgegebene interdisziplinäre Lexikon *Gedächtnis und Erinnerung* vermeidet bewusst Einträge zu den Begriffen Gedächtnis und Erinnerung selbst: „Die Vorstellung von der Funktionsweise des menschlichen Gedächtnisses ist auf Umwege angewiesen, da es selbst eine *black box* darstellt, deren Resultate und Effekte wir deutlich sehen, ohne die konkreten Vermittlungsoperationen erklären zu können.“ (Pethes & Ruchatz, 2001, S. 7). Da das Gedächtnis nicht direkt einsehbar ist, sind seine Leistungen mit vielen anderen kognitiven Tätigkeiten auf das engste und zum Teil untrennbar verknüpft. Wie Siegfried J. Schmidt schreibt: „Wer über Gedächtnis spricht, ist gezwungen, auch über Wahrnehmen und Lernen, über Wissen und Wiedererkennen, über Zeit und Erinnern zu sprechen – von Aufmerksamkeit, Emotionen und Vergessen ganz zu

schweigen.“ (Schmidt, 1991, S. 10) Umgekehrt lässt sich daraus folgern, dass das Gedächtnis im Kern dessen steht, was als Kognition bezeichnet werden kann. Dies gilt auch für den Bereich der Musikkognition. So ist es nicht verwunderlich, dass in der Einleitung zu John Slobodas für diesen Bereich grundlegendem Buch *The Musical Mind: The Cognitive Psychology of Music* das musikalische Gedächtnis als paradigmatisches Beispiel für Musikkognition überhaupt verwendet wird. Sloboda führt die bekannte Anekdote des jungen Wolfgang Amadeus Mozart an, der nach zweimaligem Hören von Allegris berühmtem *Miserere* in der Lage war, dieses Stück aus dem Gedächtnis in Noten zu transkribieren (Sloboda, 1999, S. 3f). Obwohl oder gerade weil das (Musik-)Gedächtnis auf das Engste mit der (Musik-)Kognition verbunden ist, fällt es besonders schwer, das Gedächtniskonzept in einer weitreichenden und dennoch operativ verwendbaren Definition zu fixieren.

In Kenntnis dieser ontologischen Probleme mit dem wissenschaftlichen Begriff *Gedächtnis*, setzt sich diese Arbeit nicht zum Ziel zu klären, was das Melodiegedächtnis denn nun ‚in Wahrheit‘ sei. Vielmehr soll ein Verhaltensmuster in einer sehr konkreten Situation beschrieben werden, bei welchem Menschen hypothetischerweise auf einen kognitiven Mechanismus zurückgreifen, der in der Alltagssprache *Gedächtnis* genannt wird.

Zwar ist durchaus möglich, für die wissenschaftliche Forschung einen konsistenten Gedächtnisbegriff zu operationalisieren, wie dies in psychologischen Lehrbüchern (z.B. Anderson, 1995; Edelman, 1996; Kluwe, 1996; Baddeley, 1997) und speziellen Forschungen gang und gebe ist (vgl. z.B. definitorische Diskussion bei Müllensiefen, 2000a, S. 33ff). Jedoch scheint eine spezielle Operationalisierung weder im Zusammenhang mit dem Literaturteil (Kapitel 2 und 3) noch mit den empirischen Teilen dieser Arbeit notwendig oder besonders hilfreich. Der Literaturteil arbeitet vor allem die empirische Literatur der letzten 35 Jahre auf, die sich mit dem Melodiegedächtnis beschäftigt hat. Dabei verwenden die unterschiedlichen Autoren z.T. verschiedene implizit oder explizit geäußerte operative Definitionen von Gedächtnis. Aufgearbeitet werden diese empirischen Studien jedoch nicht nach den verwendeten Definitionen des Melodiegedächtnisses, sondern anhand des methodischen Vorgehens der Forscher und anhand der Ergebnisse, die sie durch das jeweilige methodische Vorgehen erzielen. Die Auswahl der Literatur erfolgte dabei in der Regel danach, ob in der speziellen Studie von einer Beteiligung des Melodiegedächtnisses gesprochen wird oder nicht. Der Literaturteil ist damit wahrscheinlich keine Aufarbeitung aller musikpsychologischer Studien zum Melodiegedächtnis, sondern eine Aufarbeitung

einer Vielzahl von Studien, die den Terminus *Melodiegedächtnis* in der einen oder anderen Form verwenden.

Dieses Auswahlkriterium steht im Einklang mit der konstruktivistischen Interpretation empirischer Wissenschaft (vgl. Müllensiefen, 1999), nach der sich ein Forscher durch die Wahl der Methodik und die Art der wissenschaftlichen Beschreibung seinen Forschungsgegenstand selbst herstellt. Die Ergebnistabellen zum Schluss der Teilkapitel 3.2 und 3.3 formulieren zwar die empirischen Ergebnisse der Literatur zu gewissen Funktionsweisen des menschlichen Melodiegedächtnisses in knappster Form. Die jeweils vorangehenden detaillierten Beschreibungen der empirischen Forschungsmethodik zeigen aber zusätzlich, wie diese Ergebnisse hergestellt wurden und wie die „black box“ Melodiegedächtnis durch die Forscher konzeptualisiert und konstruiert wurde. Diese Form der Darstellung rüttelt nicht an der Validität und Reliabilität der empirischen Ergebnisse, d.h. es scheint i.d.R. sehr plausibel, dass unter den in den Experimenten genannten Bedingungen die menschliche Erinnerung an Melodien genau so beschrieben werden kann. Gültigkeit besitzen diese Beschreibungen aber zunächst nur unter den genannten experimentellen Bedingungen. Der Leser dieser Kapitel hat damit die Freiheit, aber auch die Bürde, selbst zu entscheiden, ob er das wissenschaftliche Konzept *Gedächtnis* zur Erklärung für das im jeweiligen Experiment erzielte Resultat verwenden will und ob er aus diesem Resultat auch Einsichten für Prozesse des Melodiegedächtnisses unter anderen Bedingungen – z.B. für eine eigene Fragestellung – generalisieren kann.

Für die empirischen Teile dieser Arbeit (Kapitel 4 und 5) ist der Umgang mit dem Begriff des Melodiegedächtnisses noch einfacher. Im Zusammenhang einer wissenschaftlichen Kooperation mit dem Institut für digitale Medientechnologie (IDMT) der Fraunhofergesellschaft an der Universität Ilmenau sollten empirische Erkenntnisse zu speziellen Forschungsfragen generiert werden. Diese Fragen hatten die Aufgabe von Forschungsleitlinien, die den Rahmen des empirischen Teils absteckten. Gleichzeitig bezogen sie sich auf einen Bereich des menschlichen Verhaltens, bei dem das Melodiegedächtnis – schon nach der alltagssprachlichen Verwendung – eine bedeutende Rolle spielt. Am IDMT wurde in den Jahren 2001 bis 2004 ein sogenanntes *Query-by-humming*-System (Qbh) für den Einsatz im Internet und auf Mobilfunkgeräten entwickelt. Qbh-Systeme haben die Aufgabe, zu einer gesummt oder gesungenen Melodie eines menschlichen Nutzers aus einer großen Melodiedatenbank die Melodie bzw. das Musikstück zu ermitteln, das der Nutzer reproduziert hat bzw. reproduzieren wollte. Qbh-Systeme können zur Ver-

kaufsunterstützung für den Musikhandel, entweder im stationären Tonträgerhandel oder in virtuellen Musikshops im Internet oder im Mobilbereich, eingesetzt werden. Anwendungen im musikpädagogischen oder im Entertainmentbereich (z.B. für Karaokemaschinen) sind ebenso denkbar.

Ein Qbh-System besteht im Prinzip aus drei technischen Modulen:

- Einem Transkriptor, der aus dem gesungenen Schallsignal eine symbolische Darstellung generiert, die im Prinzip mit der gewohnten Notendarstellung kompatibel ist.
- Einer Melodiedatenbank, die die Referenzmelodien des Systems vorhält und zu den Melodien eine Reihe von Meta-Daten (Titel, Interpret, Komponist etc. des Musikstückes, aus dem die Melodie entstammt) zuordnen kann.
- Einem Zuordnungsalgorithmus, der der gesungenen Melodie oder dem Melodieteil die Melodie(n) aus der Referenzdatenbank zuordnet, die die höchste Übereinstimmung aufweisen.

Ein Qbh-System muss, wie oben schon im Zitat von Stephen Pauws anklang, mit einer Reihe von ‚Unzulänglichkeiten‘ der menschlichen Musikverarbeitung umgehen können, um dem Nutzer sinnvolle Ergebnisse liefern zu können. Zu diesen Unzulänglichkeiten gehören vor allem weit verbreitete Schwierigkeiten bei der gesanglichen Reproduktion von Melodien (Intonationsschwierigkeiten und rhythmische Fehler) sowie die Veränderungen, denen Melodien im menschlichen Gedächtnis unterzogen werden (können). Der empirische Teil dieser Arbeit kümmert sich um die zuletzt genannten Unzulänglichkeiten des menschlichen Melodiegedächtnisses. Ziel war es, Gesetzmäßigkeiten für die Transformationen von Melodien im Gedächtnis zu finden, die innerhalb eines technischen Systems, d.h. in Form von Algorithmen und Parametereinstellungen, berücksichtigt werden können. Die im Titel dieser Arbeit genannte „Variabilität und Konstanz von Melodien in der Erinnerung“ sollte möglichst genau beschrieben werden. Da die Situation, in der eine Leistung des Melodiegedächtnisses zum Tragen kommt, innerhalb des Forschungsprojektes am IDMT relativ präzise definiert werden kann – nämlich als typische Situation, in der ein Qbh-System zum Einsatz kommt (s. Kapitel 5) – ist das wissenschaftliche Konzept von Gedächtnis für die empirische Untersuchung der Leistungen in dieser Situation zweitrangig. Der zu parametrisierende Zuordnungsalgorithmus muss nicht zwingend ein

Objekt *Gedächtnis* enthalten. Zu seiner Optimierung reichen messbare Größen für Input und Output des Qbh-Systems aus. Insofern ist eine präzise Definition von *Melodiegedächtnis* auch für den empirischen Teil dieser Arbeit nicht notwendig. Trotzdem wird in Kapitel 5 direkt von empirischen Studien ausgegangen, die angeben, Aspekte des Melodiegedächtnisses zu erforschen. Es zeigt sich, dass aus den Ergebnissen dieser Studien ein experimentelles Design abgeleitet werden kann, durch welches sinnvolle empirische Ergebnisse für die Optimierung des genannten Qbh-Systems generiert werden können. Eine genaue Definition von der Qualität des menschlichen Melodiegedächtnisses war also auch für den empirischen Teil dieser Arbeit nicht notwendig, eine Orientierung an früheren Studien zu Leistungen des Melodiegedächtnisses aber sehr ertragreich.

Die klare Leitvorstellung des Qbh-Systems war für diese Arbeit insgesamt sehr hilfreich. Denn wie der Literaturteil zeigt, ist es möglich, eine Unmenge von Einflussgrößen und Parametern im Zusammenhang mit dem Melodiegedächtnis zu untersuchen. Die ‚typische Qbh-Situation‘ hat sehr dabei geholfen, die ansonsten beliebige Wahl von Einflussgrößen und Parametern einzugrenzen und den empirischen Teil auf einen Bereich von Forschungsfragen zu konzentrieren.

1.2 Behandelte und nicht behandelte Themen dieser Arbeit

Da diese Arbeit einen Titel und einen Untertitel besitzt, die eine ganze Reihe von Assoziationen in Hinblick auf die Inhalte der Arbeit zulassen, ist es hier geboten klarzustellen, welche Themen hier behandelt und welche nur gestreift oder ganz ausgeklammert werden.

Kapitel 2 stellt drei jüngere Modelle des musikalischen (Kurzzeit-)Gedächtnisses aus der kognitiven Musikpsychologie vor, innerhalb derer auch das Gedächtnis für Melodien verortet werden kann. Die Darstellung dieser drei Gedächtnismodelle soll beispielhaft für kognitive Modelle des Musikgedächtnisses stehen und deren Geltungsanspruch und Qualität der Modellierung deutlich machen. Nicht dargestellt werden Gedächtnismodelle zu anderen auditorischen Gedächtnissystemen, wie beispielsweise dem echotischen Gedächtnis im Zusammenhang mit der sogenannten *probe-tone*-Technik (z.B. Huron & Parncutt, 1993; Leman, 2000) oder der assoziativen Organisation von Musikstücken im Langzeitgedächtnis (Halpern, 1984). Nicht dargestellt werden auch allgemeine psychologische Modelle und Theorien zu verschiedenen Gedächtnissystemen oder dem Konzept Gedächtnis überhaupt. Da im Verlauf der Arbeit dennoch des öfteren Fachtermini aus der psychologi-

schen Gedächtnisforschung verwendet werden, sei hierfür auf die einschlägigen psychologischen Lehrwerke (Kluwe, 1990; Anderson, 1995; Baddeley, 1997) oder auf die knappe Aufbereitung für den musikwissenschaftlich Vorgebildeten bei Müllensiefen (2000a, S. 32-41) verwiesen.

Das dritte Kapitel behandelt empirische Literatur zum absoluten Gehör und zu dem Gedächtnis für Einzeltöne (3.1), dem Gedächtnis für kurze Tonfolgen (3.2) und dem Gedächtnis für Melodien (3.3). Kapitel 3.2 und 3.3 referieren wahrscheinlich die Mehrzahl der veröffentlichten Studien in diesem Bereich. Die Darstellung orientiert sich dabei nicht an den bereits vorhandenen Zusammenfassungen der Literatur in diesen Forschungsfeldern (z.B. Watkins & Dyson, 1985; Nauck-Börner, 1988; Crowder, 1993; Berz, 1995; Bartlett, 1996; Vazan, 2000) und setzt sich mit den dort gezogenen Schlussfolgerungen auch nur am Rande auseinander. Vielmehr ist eine eigene Form der Darstellung gewählt, die sich besonders am methodischen Vorgehen vieler Studien und den dadurch generierten Ergebnissen orientiert und beide Bereiche kritisch diskutiert.

Im Rahmen der Methodendiskussion schien es im Kapitel 3.2 geboten, die Konstruktion von Messgrößen aus der *Signal Detection Theory* in einem kurzen Exkurs knapp aber verständlich zu erläutern, da sie zumindest in der deutschen Musikpsychologie bislang so gut wie nie detailliert behandelt wurden.

Das Kapitel 3 bietet dennoch keine systematische Methodendiskussion musikpsychologischer Forschung. Dies wäre Thema einer eigenen Veröffentlichung und wäre mit der Konzentration dieser Arbeit auf das Melodiegedächtnis nicht vereinbar.

Einige Bereiche der empirischen Literatur, die zumindest teilweise auch das Melodiegedächtnis betreffen, wurden von der Darstellung im Kapitel 3 ausgenommen, um der Fülle der Literatur insgesamt Herr zu werden und den Fokus auf die für den empirischen Teil relevante Literatur nicht zu verlieren. Außer Acht gelassen wurden erstens Befunde und praktische Methoden zum Musikergedächtnis oder zum Musiklernen, wie sie z.B. bei Winslow (1949) oder in Form der Lerntheorien sehr ausführlich von Buttram (1996) referiert werden, zweitens die Literatur zu Musikalitätstests, die in der Regel umfangreiche Testbatterien zum Melodiegedächtnis enthalten (vgl. Vispoel, 1993; Shuter-Dyson, 1999) und drittens Studien zum Gedächtnis für andere musikalische Dimensionen außer Tonhöhen, Tonfolgen und Melodien. Zu letztgenanntem Bereich gehören Studien zum Gedächtnis für Tonalitäten, Harmonien und Akkorde (z.B. Bharucha & Stoeckig, 1987; Steinke et al., 1997) oder musikalisches Timbre (Crowder, 1989; Radvansky et al., 1995; Radvansky

& Potter, 2000) und Gesangstexte in Verbindung mit Musik (z.B. Bartlett & Snelus, 1980; Serafine et al., 1984; Serafine et al., 1986; Crowder, et al., 1990; Hyman & Rubin, 1990; Wallace, 1994; Feierabend et al., 1998; Rainey & Larsen, 2002). Ebenso wenig wird in Kapitel 3 Literatur behandelt, in der bestimmte Formen der Musikkognition mit Hilfe von statistischen Strukturen modelliert werden, die aus großen Melodiesammlungen extrahiert wurden und deshalb den Begriff „memory“ verwenden. Dies trifft beispielsweise auf die Algorithmen zur melodischen Segmentierung zu, die von Bod (2001) oder Ferrand et al. (2003) entwickelt wurden.

Das vierte Kapitel widmet sich dem Thema der melodischen Ähnlichkeit. Für viele Fragestellungen im Bereich des Melodiegedächtnisses kann es von entscheidender Bedeutung sein, die Ähnlichkeit zwischen zwei Melodien genau und verlässlich bestimmen zu können. Die melodische Ähnlichkeit ist ein wichtiges Hilfsmittel z.B. bei der Analyse von aus dem Gedächtnis wiedergegebenen Melodien (Qbh-Szenario) oder der Analyse von Fehlern in Wiedererkennensexperimenten mit melodischen Materialien. Wie aber der Literaturüberblick zu Beginn von Kapitel 4 zeigt, sind zwar in der jüngeren Vergangenheit eine Vielzahl von Methoden zur Messung melodischer Ähnlichkeit vorgeschlagen worden, es fehlte bislang aber an einer vergleichenden Studie, die die unterschiedlichen Methoden vergleicht und ihre Verwendbarkeit für die Analyse von Leistungen des Melodiegedächtnisses prüft. Der Anspruch, dass Kapitel 4 diese Lücke füllen soll, erklärt und rechtfertigt seinen Umfang von etwa 120 Seiten innerhalb dieser Arbeit. Es werden nicht nur eine Vielzahl von unterschiedlichen Ansätzen zur Messung der Ähnlichkeit von Melodien für den musikwissenschaftlichen Leser nachvollziehbar dargestellt, sondern auch drei Experimente detailliert beschrieben, die zur Evaluation der verschiedenen Ähnlichkeitskonzepte dienten. Quasi als ‚Abfallprodukt‘ wird zusätzlich ein Modell der melodischen Ähnlichkeitswahrnehmung von musikalisch vorgebildeten Personen entwickelt.

Als wichtiger Ansatz für den Vergleich von Sequenzen von Symbolen, der bislang noch nicht zu den musikwissenschaftlichen Standardmethoden gehört, wird die sogenannte Edit Distance ausführlich und mit musikalischen Beispielen in Kapitel 4 erläutert.

Mit einem optimierten Algorithmus zur Bestimmung der melodischen Ähnlichkeit als Hilfsmittel wird in Kapitel 5 der Frage nachgegangen, was eine musikalisch durchschnittlich vorgebildete Person von den Melodien eines ihr unbekannten Popsongs im Gedächtnis behalten kann. Ein Wiedererkennensexperiment, das sich im Design an die Versuche von

Kauffman und Carlsen (1989) und Dowling et al. (2002) anschließt, bei der Wahl der experimentellen Parameter aber das erwähnte Qbh-Szenario mitbedenkt, soll Antworten zu dieser Frage liefern. Die Antworten haben die Form eines quantifizierten Modells, welches messbare Größen enthält, wie es als Anspruch dieser Arbeit oben formuliert wurde.

Kapitel 5 leistet sich ebenfalls einen kurzen Exkurs, in dem eine zusätzliche Messgröße zur Ähnlichkeit von Akzentstrukturen von Melodien entwickelt wird. Dieser Exkurs geht von den Akzenttheorien von Monahan et al. (1987) und Boltz und Jones (z.B. 1989) aus, wobei die dort vorgeschlagenen Akzentregeln präzise operationalisiert und zu verschiedenen Akzentmodellen kombiniert werden. Eine empirische Auswahl des besten aus den 216 definierten Akzentmodellen wurde mit Hilfe der experimentell erhobenen Daten aus Kapitel 4 geleistet.

Der zweite Teil des 5. Kapitels beschäftigt sich mit der statistischen Modellierung der Daten aus dem beschriebenen Wiedererkennensexperiment. Es werden unterschiedliche Ansätze der statistischen Modellierung beschrieben und hinsichtlich Methodik und Ergebnissen miteinander verglichen. Ein besonderes Augenmerk liegt dabei auf den sogenannten Baummodellen zur statistischen Klassifizierung und Vorhersage. Da diese Modelle bisher in der Musikpsychologie so gut wie nie zum Einsatz kamen, ist dieser Methode ein knapper, aber zum Verständnis des Prinzips und ihrer Anwendbarkeit für musikpsychologische Forschungen notwendiger Abschnitt gewidmet. Dieser Abschnitt hat – wie der zu den Messgrößen der Signal Detection Theory in Kapitel 3.2 – einen didaktischen Charakter und kann somit auch gut als Referenz für die Anwendung dieser Methode in musikpsychologischer Forschung dienen.

Obwohl diese Arbeit ein *Query-by-humming*-System als Leitvorstellung verwendet, findet sich an keiner Stelle dieser Arbeit eine detaillierte Diskussion der in der Literatur beschriebenen Qbh-Systeme oder der in ihnen implementierten Mechanismen. Abgesehen von den Algorithmen zur Bestimmung melodischer Ähnlichkeit, die ausführlich in Kapitel 4 referiert werden, wäre eine Diskussion von Qbh-Systemen in dieser Arbeit wohl eher störend als für die anvisierte Beschreibung des Melodiegedächtnisses hilfreich. Detail- und Überblicksinformationen zu unterschiedlichsten Qbh-Systemen sind in den einschlägigen Fachartikeln zu finden (z.B. McNab et al., 1996; Proceedings der ISMIR-Konferenzen 2001, 2002, 2003; Dannenberg et al., 2004).

1.3 Weitere Schritte auf diesem Forschungsansatz

Wie aus den vorangegangenen Beschreibungen der einzelnen Kapitel deutlich geworden sein sollte, erschließt diese Arbeit die einschlägige Literatur in systematischer Weise und stellt für die empirische Erforschung des Melodiegedächtnisses eine ganze Reihe von verwendbaren Konzepten und Ansätzen vor. Dabei bekommen besondere Bedeutung die Abschnitte, in denen Techniken erläutert werden, die nicht zum Standard in der Musikpsychologie gehören. Dies sind beispielsweise die Maße der Signal Detection Theory, Konzepte zur Messung melodischer Ähnlichkeit, wie sie in dem aktuellen Forschungsfeld des sogenannten Music Information Retrieval diskutiert werden (z.B. die Ansätze der sogenannten Edit Distance oder der n-grams), die Konstruktion von Messgrößen zur Bestimmung der Akzentähnlichkeit und der Vergleich verschiedener statistischer Modellierungsverfahren.

Mit diesen wissenschaftlichen Hilfsmitteln ‚an der Hand‘ wäre es möglich gewesen, noch einige Schritte weiter in Richtung quantifizierbarer Beschreibungen des Melodiegedächtnisses zu gehen. Ein interessanter Folgeschritt dieser Arbeit wäre beispielsweise die Modellierung von Leistungen des Melodiegedächtnisses mit Hilfe von Melodiedatenbanken, die die Erfahrung von Hörern widerspiegeln können.

Ein anderer möglicher und fast zwingender Schritt wäre die Überprüfung der in dieser Arbeit gewonnenen Erkenntnisse mittels eines anderen experimentellen Paradigmas, des sogenannten Wiedergabeparadigmas. Dabei müssen Versuchspersonen nicht zwei gehörte Melodien im Gedächtnis vergleichen und eine verbale Antwort zu ihrer Übereinstimmung geben, sondern sie werden gebeten, gehörte Melodien direkt aus dem Gedächtnis durch Singen oder Spielen wiederzugeben. Dieses experimentelle Paradigma hat den Vorteil, dass es dem beschriebenen Qbh-Szenario noch näher kommt und gleichzeitig einen neuen und anderen experimentellen Zugang zu Leistungen des Melodiegedächtnisses eröffnet, als es durch das Wiedererkennensparadigma möglich ist.

Diese beiden weiterführenden Perspektiven konnten leider im Rahmen dieser Arbeit nicht mehr erkundet werden. So versteht sich diese Arbeit zwar als ein wichtiger Schritt in Richtung auf eine quantifizierte Beschreibung des Melodiegedächtnisses, in jedem Fall wird sie nicht der letzte Schritt auf diesem Weg gewesen sein. Denn noch ist Guido von Arezzos eingangs zitierte Frage, warum Melodien soviel schwerer zu erinnern sind als Ackerbautechniken, nicht beantwortet.

2 Theorien des musikalischen und melodischen Gedächtnisses

Wie in der Einleitung dargelegt, richtet sich der Fokus dieser Arbeit auf quantifizierbare Modelle zur Vorhersage einzelner Leistungen des Melodiegedächtnisses bzw. der algorithmischen Beschreibung einzelner Verarbeitungsprozesse in diesem Gedächtnisbereich, wie beispielsweise die Ähnlichkeitswahrnehmung.

Die folgenden Kurzdarstellungen von Gedächtnistheorien aus der Literatur zeigen jedoch, dass bis zu diesem Zeitpunkt keinerlei Modelle des Melodiegedächtnisses existieren, die wirklich quantifizierbar sowie in realen Hörsituationen anwendbar und überprüfbar wären. Die im folgenden referierten theoretischen Modelle beschreiben zwar einzelne Komponenten und Prozesse, die das musikalische Gedächtnis oder das Melodiegedächtnis ausmachen, jedoch lassen sich eigentlich aus keinem Modell spezifische musikalische Gedächtnisleistungen unter gegebenen Bedingungen quantifizierbar ableiten. Diese Gedächtnistheorien postulieren zwar, wie ein Musik- oder Melodiegedächtnis funktioniert, sie machen jedoch so gut wie keine Aussage darüber, welche Leistungen von diesem Gedächtnissystem zu erwarten sind und wie die Gedächtnisleistungen von konkreten musikalischen Strukturen abhängen. Exemplarisch soll der Unterschied zwischen diesen beiden Ansätzen zum Erkenntnisgewinn am Beispiel der Phrasensegmentierung bei der Diskussion um Bob Snyders Theoriemodell am Ende dieses Kapitels deutlich gemacht werden.

Dennoch scheint es sehr sinnvoll, die wichtigsten vorhandenen Theorien zum Musikgedächtnis darzustellen, um für die spätere und außerhalb dieser Arbeit zu leistende Konstruktion eines quantifizierbaren Gedächtnismodells auf bereits vorhandene theoretische Konzepte oder einzelne Theoriebausteine zurückgreifen zu können.

Besondere Aufmerksamkeit soll den Theorien zum musikalischen Arbeitsgedächtnis gewidmet werden. Denn die Leistungen und Aufgaben, die im Zusammenhang dieser Arbeit von besonderem Interesse sind, dürften aufgrund des begrenzten Umfangs der zu behaltenen Objekte, der verwendeten zeitlichen Parameter und Aufgabenstellung im Experiment am ehesten dem Bereich des Arbeitsgedächtnisses zuzuordnen sein. Es ist vor allem das Modell von Berz (1995), welches auf Baddeleys Modell des Arbeitsgedächtnisses zurückgreift und auch auf den von Baddeley und Logie (1992) zusammengetragenen Erkenntnissen zum *auditory working memory* aufbaut.

Schließlich darf das Modell zum Musikgedächtnis, welches Snyder in seiner Veröffentlichung *Musical Memory* entwirft, nicht unerwähnt bleiben. Zwar ist auch Snyders Modell

nur ein hypothetisches und kein empirisch getestetes, aber Snyder leitet viele seiner Modellparameter aus empirischen Studien der Vergangenheit ab, sodass die beschriebenen Komponenten und Prozesse sehr plausibel erscheinen.

Nicht behandelt werden in diesem Kapitel theoretisch oder empirisch gewonnene Subtheorien zu einzelnen Gedächtnismechanismen, wie sie z.B. von Halpern (1984) für die assoziative Organisation von gut bekannten Musikstücken im Langzeitgedächtnis oder für Strukturreduktionen von Melodien im Gedächtnis von Polthardt-Gott (1983) oder Large et al. (1995) vorgeschlagen wurden. Hauptunterschied dieser zumindest teilweise gut quantifizierten Modelle zu den hier diskutierten Theorien ist ihre Beschränkung auf eine gewisse Klasse von Gedächtnisobjekten sowie nur wenige Gedächtnisfunktionen. Die drei im folgenden referierten Theoriemodelle liefern dagegen einen umfassenderen Rahmen für die Mehrzahl von Gedächtnisfunktionen, die im Musikgedächtnis eine Rolle spielen.

Es erscheint sinnvoll, dieses Theoriekapitel vor die Zusammenfassung der empirischen Literatur zum Ton- und Melodiegedächtnis zu setzen. Denn wie sich zeigen wird, sind viele empirische Einzelstudien nicht von größeren Gedächtnismodellen oder Theorien motiviert, sondern betrachten nur kleinere, empirisch überprüfbare Zusammenhänge zwischen einer spezifischen Gedächtnisleistung und wenigen bedingenden Variablen. Es bestünde also die Gefahr, dass sich der Leser in der Heterogenität der empirischen Studien und Einzelergebnisse zum Melodiegedächtnis verliert, wenn nicht zuvor einige Modelle bekannt sind, die die Einordnung der empirischen Einzelergebnisse dadurch erleichtern, dass sie eine Reihe von Wirkgrößen und Prozessen zumindest modellhaft miteinander in Verbindung bringen.

2.1 Modelle des musikalischen Arbeitsgedächtnisses

2.1.1 Das Modell des Arbeitsgedächtnisses nach Baddeley für melodische Inhalte

In ihrem Beitrag zu dem von Daniel Reisberg herausgegebenen Sammelband *Auditory Imagery* bringen Alan Baddeley und Robert Logie (1992) Baddeleys Modell des Arbeitsgedächtnisses mit Auditory Imagery in Verbindung. Im Ergebnis skizzieren sie jedoch anhand von experimentellen Ergebnissen, wie Prozesse des Ton- und Melodiegedächtnisses von einem Submodul des Arbeitsgedächtnisses, der sogenannten phonologischen Schleife, übernommen werden können. Sie müssen am Ende des Beitrages jedoch feststellen, dass durch die phonologische Schleife die Speicherung von musikalischen Inhalten besser erklärt wird als das Hervorrufen von auditorischen Bildern. Der Beitrag von Baddeley und Logie klärt also weniger über das Phänomen des *auditory imagery* auf als über die Fähigkeit, sich Töne und Tonfolgen zu merken, und liefert insofern für den Zusammenhang dieses Kapitels interessante Informationen.

Zunächst diskutieren Baddeley und Logie die Bedeutung des sogenannten *echoic memory* für das Phänomen des *auditory imagery*, also des bewussten Erlebens von auditorischen Vorstellungen. Sie stellen fest, dass das *echoic memory* vor allem durch das Behalten von im Umfang sehr begrenztem melodischen Material und durch sehr kurze Behaltenszeiten gekennzeichnet und zudem sehr leicht durch nachfolgende Reize zu stören ist. Das *echoic memory* kann also nicht der Sitz für die komplexen und oft sehr stabilen Leistungen der Fähigkeit des *auditory imagery* sein.

Es muss also ein ‚höheres‘ Gedächtnissystem an der Speicherung und dem willentlichen Abruf von melodischen und musikalischen Gedächtnisinhalten beteiligt sein, welches in der Lage ist, längere und komplexere Inhalte zu verarbeiten und kurzfristig vorzuhalten. Die Autoren weisen darauf hin, dass zwar einige Studien in der Vergangenheit gezeigt haben, wie Schemata und strukturelle Eigenschaften musikalischer Objekte für die Wahrnehmung und Speicherung von Musik im Langzeitgedächtnis von Bedeutung sein können (z.B. Krumhansl, 1983), dass aber für das kurzfristige Behalten von musikalischen Strukturen bisher die Mechanismen weitgehend unbekannt sind. Sie halten es aber für plausibel, dass die Mechanismen für das Behalten von verbalem und musikalischem Material sehr ähnlich oder gar identisch sind und führen als Hinweis auf diese Verwandtschaft die Ergebnisse von Salamé und Baddeley (1989) an. In jener Studie waren ähnliche Störungen des verbalen Gedächtnisses durch bestimmte Musik und verbale Störreize im Hintergrund

beobachtet worden (vgl. die ausführliche Diskussion dieser Studie in Müllensiefen, 2000a und 2000b).

Für das verbale Arbeitsgedächtnis hat Baddeley ein relativ detailliertes Modell entworfen, welches sich auf die Ergebnisse Dutzender empirischer Studien stützt. Da dieses Modell bereits in vielen Veröffentlichungen beschrieben und diskutiert worden ist, soll an dieser Stelle auf eine genauere Beschreibung des Modells verzichtet und nur eine sehr kurze, aber für das Verständnis der Studien zum Melodiegedächtnis ausreichende Zusammenfassung gegeben werden. Zum tieferen Verständnis von Baddeleys Modell sei auf Veröffentlichungen von Baddeley selbst (1986, 1997) oder anderen Autoren (Gathercole & Baddeley, 1993; Meiser, 1997; Müllensiefen 2000a, 2000b) verwiesen.

Wie in Abbildung 2-1 zu erkennen ist, ist Baddeleys Arbeitsgedächtnis hierarchisch gegliedert. An oberster Stelle steht ein System, das er zentrale Exekutive nennt und das die Zuteilung von Verarbeitungsressourcen und Aufmerksamkeit für die modalitätsspezifischen Subsysteme auf der nächst niedrigeren Hierarchiestufe regelt. Dort befinden sich zwei Subsysteme, der visuell-räumliche Skizzenblock für das visuelle und räumliche Gedächtnis und die phonologische Schleife für akustische und verbale Inhalte.

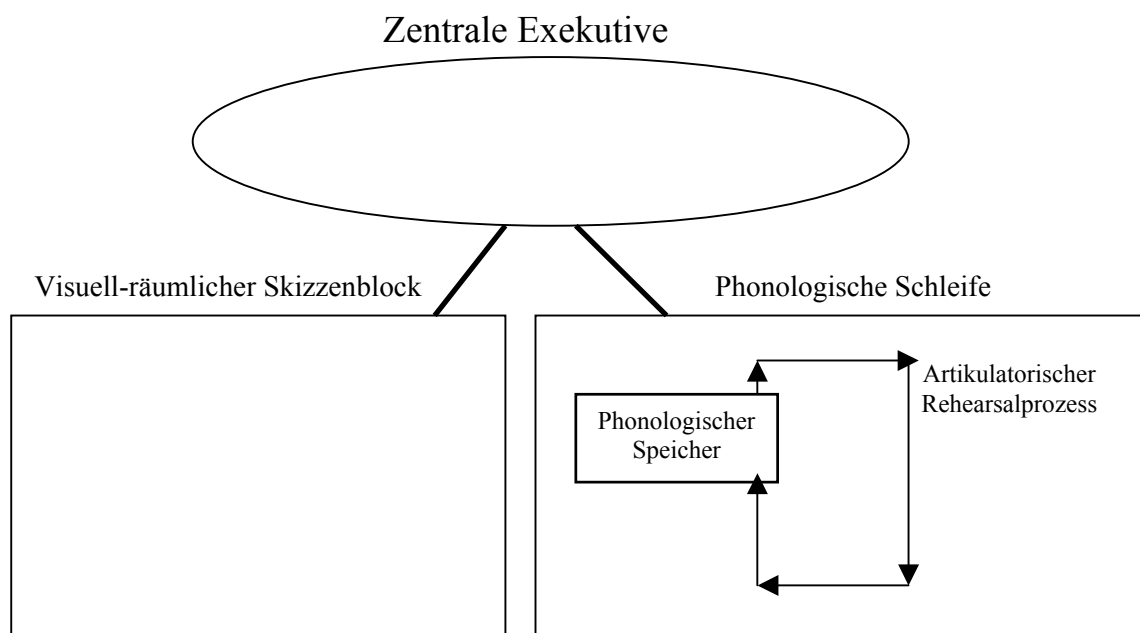


Abb. 2-1: Schematische Darstellung des Arbeitsgedächtnismodells nach Baddeley (eigene Darstellung)

Die phonologische Schleife selbst ist zweifach unterteilt, in einen artikulatorischen *Rehearsalprozess* und einen phonologischen Speicher. Müllensiefen (2000a) fasst Aufgaben und Funktionsweise des phonologischen Speichers wie folgt zusammen:

Der phonologische Speicher ist ein passiver Kurzzeitspeicher, in dem sprachliche Information während 1 ½ bis 2 Sekunden gehalten werden kann, bevor sie zerfällt. Nur akustisch dargebrachtes sprachliches Material hat direkten Zugang zu diesem Speicher, der ausschließlich auf der Grundlage phonologischer Kodierung arbeitet. Sprachliche Information kann vor dem Verlust in diesem Speicher ‚gerettet‘ werden, wenn sie vor dem Zerfall in den artikulatorischen Rehearsalprozeß übergeht, wo sie durch subvokales Wiederholen ‚aufgefrischt‘, d.h. festgehalten werden kann. Neben diesem ‚Auffrischen‘ von phonologischer Information hat der Rehearsalprozeß noch eine zweite Aufgabe. Er konvertiert oder rekodiert geschriebenes sprachliches Material in den phonologischen Kode, so dass dieses dann in den phonologischen Speicher eingehen kann. (Müllensiefen, 2000a, S. 45)

Wenn also das verbale und das musikalische Arbeitsgedächtnis auf dieselben Komponenten zurückgreifen (phonologische Schleife bzw. phonologischer Speicher und artikulatorischer *Rehearsalprozess*), dann sollten die Leistungen in kognitiven verbalen und musikalischen Aufgaben, die gleichzeitig ausgeführt werden, geringer sein als Leistungen bei gleichzeitigen musikalischen und visuellen Aufgaben, die nicht auf dieselben Ressourcen zurückgreifen. Die Identifizierung von gemeinsam beanspruchten kognitiven Systemen wird in der Regel mit dem sogenannten Zweitaufgaben-Paradigma experimentell geleistet.

Logie und Edworthy (1986) führten ein solches Zweitaufgaben-Experiment durch, in dem die Erstaufgabe in einem Wiedererkennenstest für kurze Tonfolgen bestand: Die Versuchspersonen bekamen eine tonale Tonsequenz zu hören, die sie im Gedächtnis behalten und dann mit einer zweiten Tonsequenz vergleichen mussten, die sich in einem Ton von der vorangegangenen unterscheiden konnte. Die Versuchspersonen hatten zu entscheiden, ob die beiden Tonfolgen identisch oder verschieden waren. Sie hatten diese Aufgabe alleine oder gleichzeitig eine von drei möglichen Zweitaufgaben auszuführen. Die Zweitaufgaben waren a) artikulatorische Unterdrückung (das Wiederholen einer sinnlosen Silbe wie z.B. ‚ta‘), b) ein Homophonieurteil abzugeben oder c) eine Übereinstimmung in Folgen von Sonderzeichen herzustellen. Im Ergebnis zeigte sich, dass nur die Bedingungen a) artikulatorische Unterdrückung und b) Homophonieurteile die Leistungen in der Erstaufgabe signifikant verschlechterten. In der Interpretation dieses Experiments legen Baddeley und Logie dar, dass der tonale Wiedererkennenstest offensichtlich auf zwei unterschiedliche Komponenten zurückgreift, auf eine phonologische Prozesskomponente, die auch von den Homophonieurteilen benutzt wird und auf eine phonologische Speicherkomponente, die ebenfalls die artikulatorische Unterdrückung unterstützt. Ein zweites und drittes Experiment der Studie von Logie und Edworthy (1986) stützen bzw. präzisieren diese Ergebnisse,

sodass Baddeley und Logie (1992) ein Modell zum Behalten von Tonsequenzen im Gedächtnis hypothetisch skizzieren:

Nach ihrer Modellvorstellung ist die phonologische Schleife des Arbeitsgedächtnisses ein plausibler Kandidat für die Speicherung von melodischen und wahrscheinlich auch insgesamt musikalischen Inhalten sowie auditorischen Vorstellungsbildern („auditory images“). Für das Behalten von Tonsequenzen sind beide Komponenten der phonologischen Schleife, phonologischer Speicher und artikulatorischer *Rehearsalprozess*, beteiligt. Subvokales Singen (*rehearsal*) kann für das Behalten von Tonsequenzen genutzt werden, für das Behalten von Einzeltönen ist dieser Mechanismus jedoch nicht von Nöten. Am Behalten von Einzeltönen und Tonfolgen ist in jedem Fall der phonologische Speicher beteiligt.

Baddeley und Logie vermuten also, dass das kurzfristige Behalten von verbalen und musikalischen Inhalten im Prinzip sehr ähnlich funktioniert und auf dieselben kognitiven Systeme und Ressourcen zurückgreift. Das für verbales Material gut getestete Modell des Arbeitsgedächtnisses von Baddeley sei also für die Erklärung von Gedächtnisleistungen für Tonfolgen ebenfalls anwendbar.

2.1.2 Das Modell des musikalischen Arbeitsgedächtnisses von William Berz

In einem genauso konzentrierten wie informativen Artikel skizziert William Berz – ausgehend von Baddeleys Theorieentwurf zum Arbeitsgedächtnis – ein Modell für eine separate Schleife im Arbeitsgedächtnis für musikalische oder melodische Inhalte. Er schließt sich nicht der Modellvorstellung von Baddeley und Logie (1992) an, nach der die phonologische Schleife verbale und musikalische Stimuli gleichermaßen verarbeitet, sondern postuliert eine eigene Musikgedächtnisschleife. Berz führt zwar keine eigenen empirischen Studien an, die sein Modell stützen würden, aber er referiert eine Vielzahl von Studien zum Musik- und Melodiegedächtnis, die sich letztlich unter der Annahme einer eigenen Musikgedächtnisschleife deuten lassen. Argumente dafür, dass phonologische und musikalische Schleife nicht identisch sind, findet Berz in drei Themenbereichen der empirischen Literatur.

Erstens scheint die ‚Natur‘ des Kurzzeitgedächtnisses für Musik qualitativ von der des verbalen Kurzzeitgedächtnisses verschieden zu sein. Nicht ohne einen gewissen Witz ist Berz’ Aufzählung von empirischen Ergebnissen zur Kapazität des melodischen Kurzzeitgedächtnisses. Während sich für die Kapazität des Kurzzeitgedächtnisses gemeinhin relativ starre

Grenzen etabliert haben – 7 ± 2 bzw. 5-8 Informationseinheiten (Miller, 1956) bzw. (Atkinson & Shiffrin, 1968) oder 2 Sekunden (Baddeley, 1997) – besteht in der Literatur zum Kurzzeitgedächtnis für Musik und Melodien keine Einigkeit über dessen Kapazität: Long (1977) begrenzt es auf 11-15 Töne, Pembrook (1986) auf bis zu 10 Noten. Pembrook (1987) macht eine Kapazität von 7-11 Noten von verschiedenen Faktoren abhängig, während Kauffman und Carlsen (1989) einen Kurzzeitspeicher für Musik von mindestens 180 Sekunden annehmen. Wie Deutsch (1975) und Dewitt und Crowder (1986) kommt Berz angesichts dieser widersprüchlichen empirischen Ergebnisse zu dem Schluss, dass Langzeitgedächtnisprozesse eine bedeutende Rolle in einem spezialisierten Musikgedächtnissystem spielen, welches für das kurzfristige Behalten zuständig ist. Es lässt sich interpretieren, dass aus diesem Grund die Kapazität des musikalischen Kurzzeitgedächtnisses sehr viel schwieriger zu bestimmen ist als die des Systems für verbale Inhalte.

Zum zweiten referiert Berz eine Vielzahl von Studien, die die Bedeutung von Prozesskomponenten und Strategien des Langzeitgedächtnisses für das kurzfristige Behalten von musikalischen Inhalten unterstreichen. Eine ganze Reihe von Strukturen, die auch in unbegleiteten Melodien zu finden sind, können von Versuchspersonen zur Steigerung der Behaltensleistung im Experiment eingesetzt werden. Dabei zeigt sich, dass die Ausnutzung dieser Strukturen von der Expertise bzw. dem musikalischen Vorwissen der Versuchspersonen abhängt. Die Versuchspersonen können also bereichsspezifisches Vorwissen erwerben, das ihnen effektive Wiederholungs- und Erinnerungsstrategien zur Verfügung stellt. Zu diesen musikalischen Strukturen, die durch spezifisches Vorwissen ausgenutzt werden können, zählt die Intervallstruktur (Cuddy, 1971; Cuddy & Cohen, 1976; Dewar, Cuddy & Mewhort, 1977) und die Einteilung von Melodien in sinnvolle und erinnerbare Einheiten (Long, 1977; Deutsch, 1980; Stoffer, 1985).

Insgesamt interpretiert Berz die Studien aus diesem Bereich dahingehend, dass das musikalische Arbeitsgedächtnis eine wichtige Prozesskomponente besitzt, die analog zum artikulatorischen *Rehearsalprozess* in der phonologischen Schleife arbeiten mag, in der aber musikspezifische Strategien benutzt werden können.

Als Drittes führt Berz Studien zur Wirkung von Hintergrundmusik auf verschiedene gleichzeitige kognitive Tätigkeiten auf. Ähnlich wie Müllensiefen (2000a) konstatiert auch Berz, dass Hintergrundmusik keine einheitliche Wirkung hat, sondern dass Wirkung von Hintergrundmusik offensichtlich von einem komplexen Zusammenspiel von kognitiver Primäranforderung und Eigenschaften der Musik abhängt. In jedem Fall konsistent scheint

die Störwirkung von Vokalmusik im Gegensatz zu Instrumentalmusik auf verbale Tätigkeiten größer zu sein (Martin, Wogalter & Forlano, 1988; Salamé & Baddeley, 1989). Auch dies scheint nach Berz für ein eigenes Subsystem des Arbeitsgedächtnisses für Musik zu sprechen.

Berz stellt nach der Durchsicht der Studien innerhalb der genannten drei Bereiche fest, dass kurzfristige musikalische Gedächtnisleistungen nicht vollständig mit der phonologischen Schleife aus Baddeleys *Working Memory* erklärt werden können, dass das eigene Musikgedächtnis aber strukturell große Ähnlichkeit zur phonologischen Schleife haben mag. Sein Modell, welches in Abbildung 2-2 dargestellt ist, ist demnach an Baddeleys Grundmodell angelehnt.

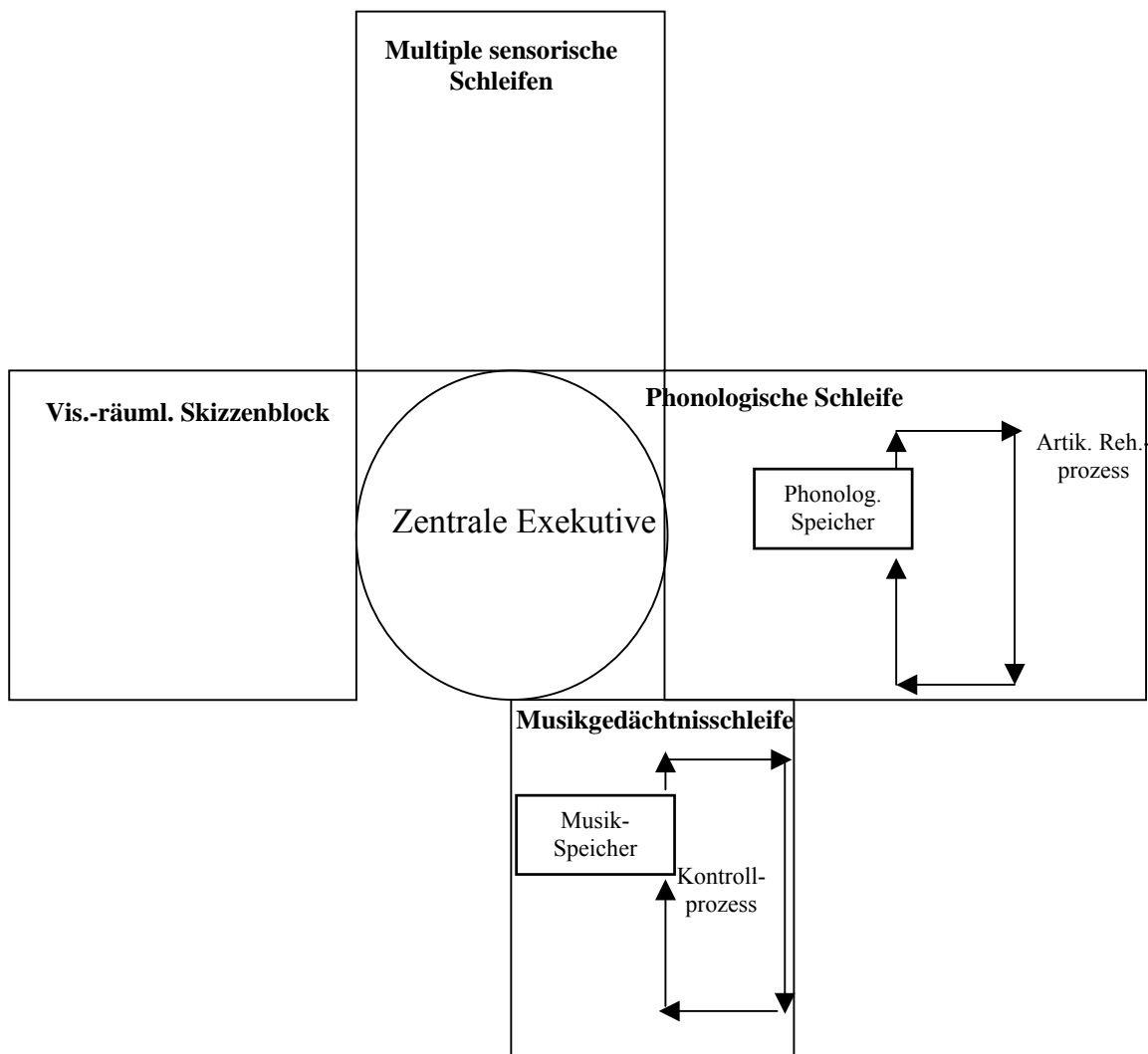


Abb. 2-2: Modell des Arbeitsgedächtnisses inklusive des musikalischen Arbeitsgedächtnisses nach Berz; eigene Darstellung, angelehnt an Berz (1995, S. 362)

Im Vergleich mit dem in Abb. 2-1 dargestellten Originalmodell nach Baddeley kommt eine Musikgedächtnisschleife hinzu, die mit der zentralen Exekutive verbunden ist und gleichzeitig eine lose Beziehung zur phonologischen Schleife hat. Weiterhin sind auch weitere Schleifen für die übrigen Sinnesmodalitäten (Geruch, Geschmack etc.) vorgesehen, wie Berz vermutet.

Die angenommene musikalische Gedächtnisschleife ist, genau wie die phonologische Schleife, zweigeteilt, und zwar in einen musikalischen Speicher und einen eigenen Kontrollprozess, der auf innerer musikalischer Sprache beruht, die Berz mit dem von Gordon (1986) geprägten Konzept der ‚Audiation‘ vergleicht. Leider beschreibt Berz die Charakteristiken der beiden Komponenten nicht näher, sondern stellt sie nur in Analogie zu den Bestandteilen der phonologischen Schleife. Er unterstreicht jedoch, dass Speicher- und Prozesskomponenten bei bestimmten Aufgaben, wie dem Musikediktat, im Wettbewerb stehen können. Unterschiede in Kapazität und Leistungsfähigkeit des musikalischen Arbeitsgedächtnisses mögen in der Regel auf erworbene Strategien des musikalischen Langzeitgedächtnisses zurückgehen (verbale Rekodierung, Verarbeitung der musikalischen Syntax). Diese Unterschiede sollten beispielsweise in den verschiedenen Leistungen von Personen mit unterschiedlicher musikalischer Vorbildung bei der Bewältigung von Musikalitätstests ihren Niederschlag finden.

2.2 Übergreifende Modelle des musikalischen Gedächtnisses

2.2.1 Das Drei-Zeiten-Modell von Bob Snyder

Die Leistung von Bob Snyders Buch *Music and Memory. An Introduction* (2000) liegt vor allem in der Aufbereitung und Darstellung aller möglichen Aspekte der kognitiven Musikverarbeitung, die mit Gedächtnisprozessen in Verbindung gebracht werden können. So referiert er beispielsweise die Rolle kognitiver Konzepte wie Kategorien (Kap. 7), Schemata (Kap. 8) und der Metapher (Kap. 9), genauso wie Forschungsergebnisse zu musikalischen Konzepten wie Melodie (Kap. 11), Rhythmus (Kap. 12) oder Form (13). Snyders Buch wird damit zum „excellent textbook for an introductory course in the psychology of music“, wie Roger Chaffin in seiner Rezension in *Music Perception* schreibt (Chaffin, 2002, S. 106).

Wenig bis gar nichts Verwertbares enthält das Buch in Hinsicht auf den im Verlauf dieser Arbeit entwickelten empirischen Ansatz der Quantifizierung von Gedächtnisleistungen bzw. deren Vorhersage anhand einer gegebenen Experimentalsituation. Das Buch stellt

sehr gut die möglichen und komplexen Zusammenhänge zwischen verschiedenen musikalischen Strukturen und kognitiven Verarbeitungsprozessen dar. Snyder stellt jedoch nur an wenigen Stellen Modelle auf, die wirklich falsifizierbar und damit prüfbar wären. Dennoch werden einige interessante Zusammenhänge hergestellt, aus denen sich im Nachhinein empirisch prüfbare und vielversprechende Modelle entwickeln ließen. Dies trifft z.B. auf die Verwendung von kognitiven Schemata bei der Musikverarbeitung oder die Angaben zum echotischen Gedächtnis und der frühen Verarbeitung zu (Kap. 2).

Von besonderem Interesse im Zusammenhang dieses Kapitels ist das hypothetische Gedächtnismodell, das Snyder gleich im ersten Kapitel darstellt und auf das er an zahlreichen Stellen im weiteren Verlauf des Buches rekurriert. Im Gegensatz zu den beiden zuvor referierten Modellen beschränkt sich sein Modell nicht auf das kurzfristige Behalten von musikalischen Inhalten, sondern erhebt den Anspruch, alle möglichen Gedächtnisprozesse zu umfassen.

Snyders Gedächtnismodell ist dreigeteilt, und zwar entlang einer Zeitachse in unterschiedliche Verarbeitungsstufen (Snyder, 2000, S. 12ff). Auf diesen drei Ebenen der Verarbeitung wird das physikalische Kontinuum, das Musik in der Regel darstellt, in unterschiedliche psychische und sprachliche Objekte eingeteilt. Die drei Verarbeitungsstufen korrespondieren in Synders Modellvorstellung eng mit verschiedenen Arten der Informationsverarbeitung. Zur Darstellung der Verarbeitungsprozesse verwendet er entsprechend auch Beschreibungen aus der Psychophysik, der Gestaltpsychologie und der kognitiven Psychologie. Erkenntnisse aus diesen drei Theoriebereichen orientieren und stützen sein Gedächtnismodell. Zusätzlich werden auch einige Erkenntnisse zur neuronalen Verarbeitung von Musik eingeflochten. Die Dreiteilung des Modells und das Verbinden von musikalischen, psychologischen und neuronalen Strukturen auf verschiedenen Ebenen ist – wie Snyder anführt – von Ideen und Vorstellungen Karlheinz Stockhausens (1989) inspiriert, der bereits ähnliche Analogien herstellte.

Abbildung 2-3 zeigt Bob Snyders Modell im schematischen Überblick. Eine kurze Zusammenfassung der Beschreibungen der Verarbeitungsstufen soll einen Überblick über das Modell geben.

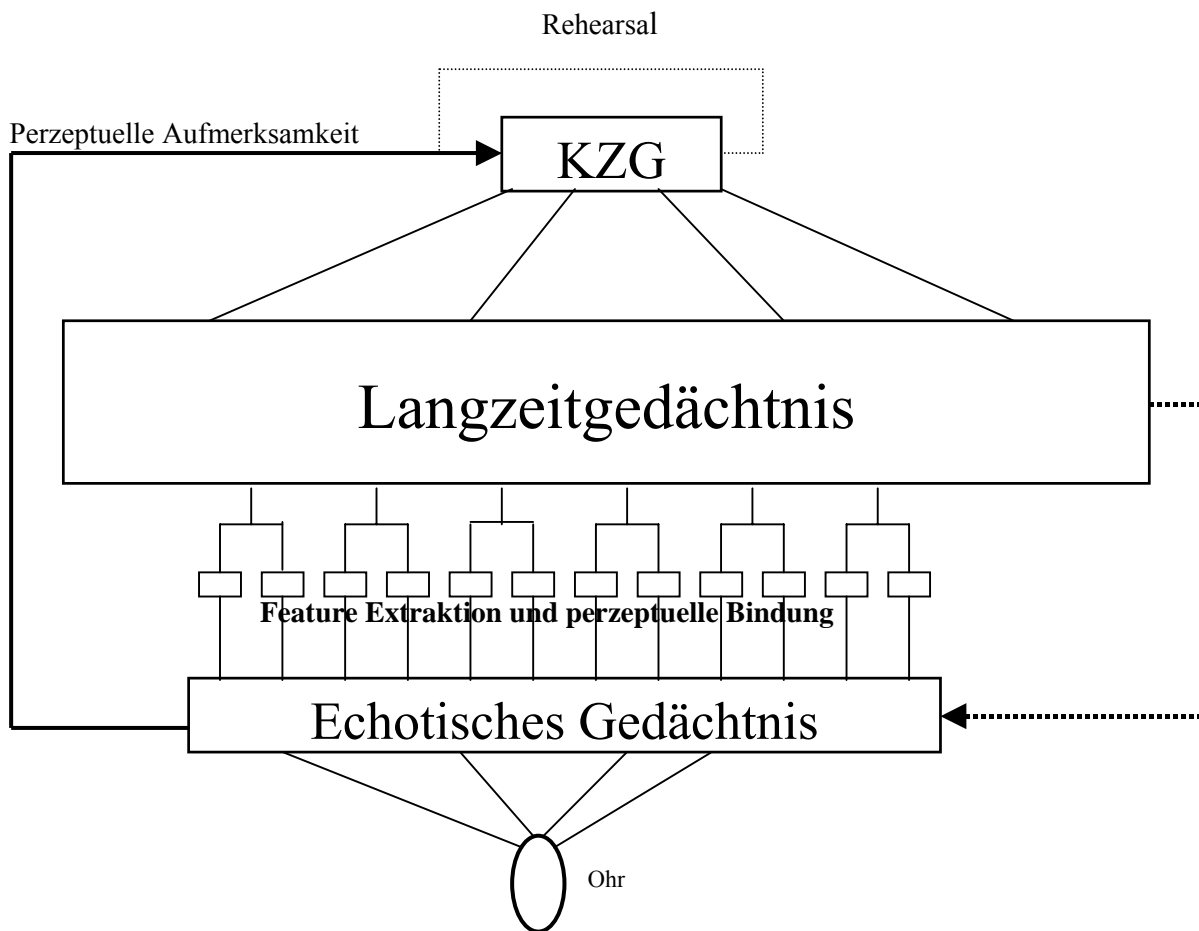


Abb. 2-3 Schematische Darstellung des auditorischen Gedächtnisses nach Snyder (2001); eigene Darstellung nach Snyder (2001, S. 6).

Die unterste Zeitstufe nennt Snyder den „Event Fusion Level“. Dieser entspricht dem aus der Psychologie bekannten echotischen Gedächtnis. Es werden Zeitfenster von weniger als 50 Millisekunden verarbeitet. Insgesamt wird sehr viel auditorische Information für sehr kurze Zeit vorgehalten. In Anlehnung an Crowder (1993) geht Snyder von einer Kapazität von 250 Millisekunden bis wenigen Sekunden aus. Auf diesem Level können einzelne Eigenschaften aus dem Soundsignal entnommen werden („Feature Extraction“). Zu diesen Eigenschaften gehören die Tonhöhe von periodischen Signalen, aber auch Timbre und Lautstärkeinformationen und deren Änderung in der Zeit. Der Input dieses Gedächtnissubsystems sind die Impulse der Nerven des Innenohres. Die Extraktion oder Konstruktion von Eigenschaften aus dem Soundsignal findet dann in den nachfolgenden neuronalen Zentren statt. Hier wird die Information des kontinuierlichen Soundsignals kategorisiert, obwohl Snyder zugesteht, dass auch Überbleibsel der kontinuierlichen auditorischen Information auf höheren Verarbeitungsstufen bewusst wahrgenommen werden können, z.B. als Nuancen (in Klangfarbe, Tonhöhe etc.). Die Bindung von einzelnen Eigenschaften ist

eine elementare Assoziation von zeitlich benachbarten Elementarereignissen. So werden hier z.B. Grund- und Obertöne zu einer Einheit verschmolzen. Die Prozesse auf dieser Verarbeitungsstufe werden von den eingehenden akustischen ‚Rohdaten‘ beherrscht, und es findet eine sogenannte *Bottom-Up*-Verarbeitung statt, die unbewusst abläuft.

Snyder führt noch einige Ergebnisse psychoakustischer Forschung an, die diese Ebene charakterisieren. Dazu gehören die zeitlichen Grenzwerte zur Verarbeitung von akustischen Ereignissen in sehr kurzen Abständen, die generell alle mit der Verarbeitung auf niederen neuronalen Verarbeitungsstufen in Verbindung gebracht werden. Insgesamt setzt er auf dieser Ebene die frühe neuronale Verarbeitung mit der psychoakustischen Ereignisverschmelzung und dem echotischen Gedächtnis gleich.

Auf der nächsten zeitlichen Verarbeitungsstufe werden rhythmische und melodische Information gruppiert und in sinnvolle Einheiten zusammengestellt. Diese Zeitstufe reicht von 1/16 Sekunde (62,5 Millisekunden) bis zu acht Sekunden. Im unteren Zeitbereich dieser Verarbeitungsstufe spielen vor allem die sogenannten Gruppierungsmechanismen eine bedeutende Rolle. Snyder unterscheidet verschiedene primitive Gruppierungsmechanismen, die angeboren und der menschlichen Art insgesamt eigen sind. Dazu gehören das Zusammenfassen von mehreren akustischen Ereignissen durch irgendeine Art von Abschluss („closure“) und die Tendenz, akustische Ereignisse bestimmten Eigenschaften einer Schallquelle zuzuordnen. Diese Arten der Gruppierung sind unabhängig vom psychologischen Konzept des Kurzzeitgedächtnisses.

Das Kurzzeitgedächtnis kommt jedoch anschließend ins Spiel, bei der Wahrnehmung von melodischen oder rhythmischen Motiven oder Zellen, die die kleinste Einheit musikalischer sinntragender Elemente sind, die mit dem Kurzzeitgedächtnis verarbeitet wird. Diese Zellen enthalten nach Snyder maximal fünf Elemente. Die Grenzen dieser Einheiten können anhand der bekannten Galtsgesetze – Nähe, Ähnlichkeit, gute Fortführung – ausgemacht werden.

Ausgedehnter als diese Motive oder Zellen sind die musikalischen Phrasen, die die größte Einheit darstellen, die das musikalische Kurzzeitgedächtnis verarbeiten kann. Die Länge von Phrasen wird generell durch die Atemperioden des Menschen oder durch andere physiologische Gegebenheiten begrenzt. Auch die gemeinhin angenommene Kapazitätsgrenze des Kurzzeitgedächtnisses von wenigen Sekunden setzt musikalischen Phrasen eine Grenze. Dadurch können Phrasen jedoch vollkommen in der psychischen Gegenwart wahrgenommen werden. Auch für die Einteilung in Phrasen spielen nach Snyder die drei

genannten Gestaltgesetze die wichtigste Rolle, wobei er die Nähe, vor allem die zeitliche Nähe, für den wichtigsten Gruppierungsfaktor hält.

Die Stufe der melodischen und rhythmischen Gruppierung korrespondiert nach Snyder mit dem seit Atkinson und Shiffrin (1968) etablierten Konzept des Kurzzeitgedächtnisses (*short-term memory*). Ein bedeutender Unterschied dieser Stufe zum *Event Fusion Level* ist das Zusammenfassen von Ereignissen zu Gruppen. Es werden nicht Grenzen zwischen Einzelereignissen wahrgenommen (wie z.B. die Grenzen zwischen unterschiedlichen Tönen), sondern es werden mehrere Ereignisse zu Gruppen zusammengefasst. Im Unterschied zu den beiden anderen Verarbeitungsebenen hinterlässt die Verarbeitung auf dieser mittleren Ebene keine anatomischen oder chemischen Änderungen in den beteiligten neurologischen Strukturen. In Rückgriff auf den sogenannten Papez-Kreis (Papez, 1937) nimmt Snyder einen kreisförmigen Strom von neuronaler Aktivität an, der die Inhalte des Kurzzeitgedächtnisses erhält.

Zur Kapazität und Begrenzung dieses Gedächtnissystems führt Snyder die ‚unvermeidlichen‘ 7 ± 2 Elemente an, die auf Millers bekannten Aufsatz zurückgehen (Miller, 1956). Snyder ergänzt diese starren Gedächtnisgrenzen jedoch durch neuere Erkenntnis zum *chunking* (d.h. zur sinnvollen Organisation von Gedächtnisinhalten in festen Gruppen) und zur Verwendung von anderen Gedächtnisstrategien zur Kapazitätserweiterung. So hält er es für möglich, dass durch *chunking* auf fünf möglichen Hierarchieebenen die Anzahl der speicherbaren Elemente auf 5^5 (also etwa 3000) erweitert werden kann. Gerade im Bereich der Musik seien das *chunking*, welches auf Vorwissen aus dem Langzeitgedächtnis zurückgreife und die musikalischen Assoziationen sehr leistungsfähige Mechanismen.

Als letzte Verarbeitungsstufe schließt sich in Snyders Modell die Ebene der Form an. Diese verarbeitet musikalische Geschehnisse in einem Zeitfenster von ca. 16 Sekunden bis zu einer Stunde. Es werden hier ganze Abschnitte von Musik verarbeitet. Die verarbeiteten Ereignisse liegen in der psychischen Vergangenheit und können zeitlich geordnet werden; sie werden also nicht sofort bei der Verarbeitung, sondern in einer Rückschau auf das Gehörte in Beziehung gesetzt. Diese Rückschau findet in dem ebenfalls von Atkinson und Shiffrin populär gemachten Langzeitgedächtnis (LZG) statt. Aus den Ereignissen auf dieser Formebene werden in der Wahrnehmung des Hörers ganze Musikstücke durch Assoziation zusammengesetzt. Die Aktivierung eines Elementes im Langzeitgedächtnis kann – je nach Stärke der Assoziation – andere Elemente ganz oder halb aktivieren. Der zugehörige und gedächtnispsychologisch gut untersuchte Mechanismus ist das sogenannte *priming*

(vgl. Grote, 2001), welches durch Vorgabe von speziellen Hinweisreizen die Verarbeitung von bestimmten Gedächtnisinhalten erleichtert.

Snyder weist darauf hin, dass die Art der Kodierung von musikalischen Ereignissen im Langzeitgedächtnis eine übergeordnete Bedeutung für den Abruf von Inhalten aus dem Langzeitgedächtnis hat. Er unterstreicht weiterhin richtigerweise, dass Gedächtnisinhalte keine direkte Repräsentation von physikalischen Erfahrungen seien, sondern durch kognitive Kontexte und Erfahrungshorizonte mitgeformt werden. Folgerichtig schließt er einige Kapitel an seine Beschreibung des musikalischen Langzeitgedächtnisses an, die sich mit verschiedenen Mechanismen des LZGs beschäftigen, so z.B. mit der kognitiven Kategorienbildung, dem Ansatz der kognitiven Schemata oder Skripte und der musikalischen Metapher.

Als Beispiel für die Art der wissenschaftlichen Ergebnisse und der Methode Sydners in seiner gesamten Monographie kann man seine Beschreibung von musikalischen Phrasen verwenden. Hieran lassen sich gut die Unterschiede zwischen Sydners qualitativer Modellbildung und anderen, eher empirisch orientierten Modellen zeigen. Die Phrasensegmentierung ist ein besonders gutes Beispiel, da für diesen Verarbeitungsprozess eine Reihe von weiteren Modellen verschiedenster Autoren bereits veröffentlicht sind, und da sie in den empirischen Kapiteln (Kap. 4 und Kap. 5) dieser Arbeit noch von Bedeutung sein wird.

Sydners Methode besteht vor allem in der Diskussion, Interpretation und plausiblen Verbindung von Forschungsergebnissen zu diesem Bereich. Der Mechanismus zur Phrasensegmentierung, den er beschreibt, gibt seine Einschätzung der wahrscheinlichsten oder sinnvollsten Wirkmechanismen wider. Er vermeidet aber eine operationalisierbare Definition von Regeln, die seine Vorstellung der Phrasensegmentierung ohne weiteres anwendbar machen würden. Im Gegensatz dazu sind beispielsweise von David Temperley (2001) und Emiliós Cambouropoulos (2003) weitgehend exakt definierte Algorithmen veröffentlicht worden, die eine empirische Prüfung an konkreten Musikstücken erlauben. Das Beispiel der Phrasensegmentierung steht paradigmatisch für Sydners Vorgehen in fast allen anderen Bereichen seines Buches, plausible und kohärente Ideen zu Verarbeitungsmechanismen ausgehend von der existenten Literatur zu entwerfen, die eigentliche Operationalisierung und Modellbildung aber anderen Forschern zu überlassen. Dies soll nicht bedeuten, dass sein Vorgehen bei der Theoriebildung im Bereich der Musikgedächtnisforschung oder im Kontext dieser Arbeit weniger wertvoll wäre. Ganz im Gegenteil, ist die Formung von kohärenten Ideen zur Modellbildung aus den Erkenntnissen der bestehenden Literatur von

ganz basaler Bedeutung. Ohne eine kohärente Vorstellung von einem Gesamtsystem Musikgedächtnis wäre eine exaktere Modellbildung in einzelnen Teilbereichen wahrscheinlich gar nicht möglich oder würde wenig sinnvolle und widersprüchliche Ergebnisse zeitigen. Zusätzlich wird erst durch diese Ebene der Modellbildung eine Generalisierung von einzelnen empirischen Forschungsergebnissen möglich, zulässig oder auch unzulässig.

Es sollte jedoch auch klar sein, dass der Fokus dieser Arbeit auf einer möglichst genauen Operationalisierung von spezifischen Gedächtnisleistungen und eng umgrenzten Bedingungen liegt, die eine empirische Prüfung und eine Optimierung des Modells erst ermöglichen. Die Ergebnisse und das Modell, das Snyder in seinem Buch referiert, sind für eine Parametrisierung eines empirischen Gedächtnismodells nicht ausreichend. Unter anderem macht gerade diese Lücke, die Snyder in seiner Monographie lässt, die vorliegende Arbeit sinnvoll und notwendig.

Obwohl Titel und Klappentext von Snyders Veröffentlichung suggerieren könnten, dass es sich um eine erschöpfende Darstellung des menschlichen Musikgedächtnisses handle, bleiben dort einige Fragen, Zusammenhänge und Details nur angedeutet, die im Fortgang dieser Arbeit genauer beleuchtet werden sollen. Die Kapitel 3, 4 und 5 dieser Arbeit könnte man somit als Fortsetzung oder Ergänzung zu Snyders theoretischen und an qualitativen Modellen orientierten Ansatz verstehen.

3 Das Gedächtnis für Töne, kurze Tonfolgen und Melodien

(Literaturübersicht)

3.1 Gedächtnis für Einzeltöne

Da diese Arbeit sich mit dem Gedächtnis für Melodien und Tonfolgen beschäftigt, ist eine Behandlung des Gedächtnisses für Einzeltöne zwar nicht von zentraler Bedeutung, kann aber auch nicht völlig vernachlässigt werden. Denn gerade die Beziehungen und Unterschiede zwischen Leistungen und Funktionsweisen eines Tongedächtnisses und den Gedächtnisleistungen für Melodien, können Aufschluss darüber geben, was das Melodiegedächtnis als spezifisch musikalisch charakterisiert.

Nur in aller Kürze soll im folgenden auf das absolute Gehör eingegangen werden, welches als besonderes Phänomen in der Vergangenheit schon vielfach Gegenstand musik- und neuropsychologischer Forschung war und entsprechend detailliert in vielen Lexikoneinträgen und Spezialveröffentlichungen behandelt wurde (vgl. z.B. Rakowski & Morawska-Büngeler, 1987; Heyde, 1987; Altenmüller in MGG Sachteil; Takeuchi & Hulse, 1993; Chin, 1997; Ward, 1999; Zatorre, 2003; Vitouch, 2004).

Interessanter im Zusammenhang dieser Arbeit sind dagegen einige Forschungen, die sich mit dem Gedächtnis für absolute Tonhöhen bei Nicht-Absoluthörern beschäftigen. Auch in Hinsicht auf die Frage, zu welcher Art von Erinnerungsleistungen die große Mehrheit der potentiellen Nutzer eines *Query-by-humming*-Systems fähig ist, sind gerade die Ergebnisse dieser Studien von Bedeutung. Wäre beispielsweise vorauszusehen, dass sich gesungene Wiedergaben von Melodien aus dem Gedächtnis mit einer gewissen Wahrscheinlichkeit in einem geringen Bereich um die Originaltonhöhe herum bewegen, so könnte dies zu einer Optimierung der Heuristiken und Algorithmen genutzt werden, die den Abgleich einer menschlichen *query* mit einer Melodiedatenbank leisten.

Vor allem um den Begriff des absoluten Gehörs vom Gedächtnis für absolute Tonhöhen abzugrenzen, soll zunächst überblickartig der Terminus ‚absolutes Gehör‘ erläutert werden. Im darauf folgenden Abschnitt werden dann diejenigen Studien kurz referiert, die das Behalten von absoluten Tonhöhen primär unter einem Gedächtnisaspekt betrachten.

3.1.1 Absolutes Gehör

Die folgenden Ausführungen über das absolute Gehör (AG) referieren einige der Aspekte, die in den umfangreichen Überblicksveröffentlichungen von Heyde (1987), Ward (1999)

und Vitouch (2004) z.T. sehr ausführlich dargestellt werden. Da sie im Zusammenhang dieser Arbeit von sehr untergeordneter Bedeutung sind, werden einige Aspekte des AGs, die üblicherweise in ausführlicheren Abhandlungen zum AG zu finden sind, hier ausgespart. Dazu gehören die neurologische Basis des AGs, die Nützlichkeit der Fähigkeit für das Musizieren und der Zusammenhang zwischen AG und musikalischer (Hoch-)Begabung.

Die Definitionen des absoluten Gehörs in der Literatur sind relativ homogen. Es besteht in der Regel Einigkeit darüber, was das Konzept ‚absolutes Gehör‘ bedeutet. Eine knappe, gut verständliche Definition in deutscher Sprache liefert beispielsweise Altenmüller im entsprechenden Abschnitt des Eintrages Gehör in der neuen MGG (Spalte 1101-1102):

Als absolutes Gehör wird die Fähigkeit bezeichnet, die Tonhöhe isolierter Töne ohne äußere Referenz benennen oder produzieren zu können.

Ganz ähnlich die Definition von Ward (1999, S. 265) auf englisch:

The ability to identify the frequency or musical name of a specific tone, or, conversely, the ability to produce some designated frequency, frequency level, or musical pitch without comparing the tone with any objective reference tone (i.e., without using relative pitch [[RP]]).

Wie Heyde (1987) und Vitouch (2004) darstellen, reicht die Forschungstradition zum AG über 100 Jahre zurück (Begriff von Carl Stumpf (1883) eingeführt, erste Monographie von Otto Abraham (1901)) und die seitdem erschienenen Veröffentlichungen zu diesem Phänomen sind Legion (gemessen an anderen Themengebieten der Musikpsychologie). Bis zum gegenwärtigen Zeitpunkt dürften knapp 400 Studien zum AG erschienen sein.

Mehrere Aspekte des AGs haben wahrscheinlich das übergroße Forschungsinteresse an diesem Phänomen begründet, und großteils sind es genau diese Aspekte, die es auch im Vergleich mit dem Tongedächtnis von Normalhörern im Zusammenhang dieser Arbeit interessant machen. Zu diesen Aspekten gehören:

- Die Seltenheit des AGs.
- Die Theorien zum Erwerb des AGs bzw. die Frage, ob es erlernt werden kann.
- Die differenzierten Untersuchungstechniken zur genauen Bestimmung des Typs und der Ausprägung sowie die zugehörigen Modellvorstellungen des AGs.

Zu diesen drei Aspekten sollen in je einem kurzen Abschnitt die Erkenntnisse aus der Literatur referiert werden.

Das Vorkommen des absoluten Gehörs in der Bevölkerung westlicher Kulturen wird allgemein zwischen ca. 1:10.000 und 1:1.000 (Bachem, 1955; Profita & Bidder, 1988; Takeuchi & Hulse, 1993) geschätzt. In Teilpopulationen, wie bspw. bei professionellen Musikern ist das Vorkommen jedoch deutlich häufiger und wird auf 1:100 bis 1:5 geschätzt (z.B. Miyazaki, 1988; Baharloo et al. 1998). Das AG ist in jedem Fall eine seltene Fähigkeit, die unter Musikern deutlich häufiger anzutreffen ist. Eine auffällige Häufung findet sich unter nahen Verwandten. Die Wahrscheinlichkeit dafür, dass eine Person ein absolutes Gehör besitzt, ist 8-15fach höher, wenn sich unter den Geschwistern dieser Personen mindestens ein weiterer Absolut Hörer befindet. Die Schätzung der Prävalenz in der Gesamt- bzw. einer musikalisch vorgebildeten Teilpopulation hängt jedoch wesentlich von der Erhebungsmethode (Selbsteinschätzung vs. Test) und der Art des Testverfahrens ab (z.B. Toleranz von Oktav- und Halbtonfehlern).

Die Seltenheit des AGs insgesamt und das vermehrte Auftreten unter nahen Verwandten haben schon früh den Verdacht entstehen lassen, dass es sich beim AG um eine genetisch vererbte Fähigkeit bzw. Disposition zur Ausbildung dieser Fähigkeit handelt. Im Anschluss an Ward und Burns (1982) teilt Vitouch (2004) die Erklärungsmodelle zum Erwerb des AG in vier unterschiedliche Ansätze: Die Vererbungstheorie, die Prägungstheorie (in einer frühkindlichen Phase), die Erlernungstheorie und die Verlernungstheorie. Obwohl sich bislang kein Erklärungsansatz mit letztendlicher Sicherheit gegenüber den anderen durchgesetzt hat, neigen Ward (1999) und Vitouch – wie Eva-Marie Heyde und viele andere Forscher auch – der Prägungstheorie zu. Die Sensibilisierung von Kindern durch Lernerfahrungen bis zu einem Alter von fünf bis sechs Jahren, bei denen die Bezeichnung von absoluten Tonhöhen von Bedeutung ist oder das musikalische Gehör besonders geschult wird, scheint jedoch eine zumindest notwendige Bedingung zu sein. Eine interessante Theorie, die das absolute Gehör mit der Aussprache von Wörtern auf einer absoluten Tonhöhe experimentell vergleicht und die Ergebnisse mit dem Erlernen einer Tonsprache im Kindesalter in Beziehung setzt, ist vor kurzem von Deutsch und Kollegen veröffentlicht worden (Deutsch et al., 2004). Das problemlose Erlernen sowohl des absoluten Gehörs wie der Tonregeln im Mandarin oder im Vietnamesischen sei demnach vor allem in einer sehr frühen Phase möglich. Da beide Fähigkeiten auf dieselben Gehirnstrukturen zurückgriffen (Deutsch et al., 2004, S.352), müsse das absolute Gehör bei Sprechern einer Tonsprache deutlich weiter verbreitet sein. Ein epidemiologischer Nachweis dieser Annahme fehlt jedoch bislang. Gerade die neuropsychologische Forschung in diesem Bereich neigt aber zu

der Annahme, dass Personen asiatischer Abstammung generell eine günstige genetische Disposition für die Ausbildung des AGs mitbringen (vgl. Zatorre, 2003, S. 694).

Weder von Deutsch und Kollegen, noch von anderen Autoren ist die Frage, ob eine besondere (und seltene) genetische Prädisposition zusätzlich vorhanden sein muss, oder ob es eher Faktoren des Umfeldes sind, die in einer ‚günstigen‘ Weise wirken müssen, bis jetzt nicht endgültig beantwortet worden. Ein gewichtiges Argument für die Bedeutung des musikalischen Umfeldes zur ‚richtigen Zeit‘ ist der starke Zusammenhang zwischen dem Alter, in dem das aktive Musikmachen begonnen wird und dem Auftreten von AG im Erwachsenenalter. Nach Erhebungen von Miyazaki (1988) und Gregersen et al. (1999) besitzen bis zu 50% der betrachteten Population von japanischen Musikern das AG. Da die japanische Musikerziehung in der Regel sehr früh beginnt und wie im Beispiel der Suzuki-Methode sehr am Gehör orientiert ist, spräche dies gegen eine besondere oder seltene Prädisposition. Auch eine Befragung von britischen Profimusikern (Sergeant, 1969) zeigte den Zusammenhang zwischen dem Beginn des Instrumentalspiels und dem späteren Besitz von AG auf: 88% der Gruppe von Musikern, die im Mittel mit 5,6 Jahren mit dem Instrumentalspiel angefangen hatten, besaßen AG, dagegen waren es in der Gruppe mit mittleren Einstiegsalter von 9,9 Jahren 0%.

Eine echte Validierung dieses Zusammenhangs fehlt leider jedoch in beiden Studien, da nur Musiker befragt wurden und weil jene Personen, die zwar auch frühzeitig Instrumentalunterricht genossen, aber später keine musikalische Karriere verfolgt hatten, bisher in keiner Studie als Vergleichsgruppe befragt oder getestet werden konnten. Auch Untersuchungen mit systematischem Training von Kindern erbrachten hier keine eindeutigen (Crozier, 1997; Sakakibara, 1999, 2000) oder verallgemeinerbaren Ergebnisse (vgl. die Beschreibung von Wards Experiment mit seinen eigenen vier Töchtern, 1999, S. 293). Als gesichert kann jedenfalls gelten, dass das Training im Erwachsenenalter, wenn überhaupt, nur unter Einsatz von viel Zeit und Motivation eine Fähigkeit hervorbringen kann, die annähernd mit den Leistungen früh geprägter Absoluthörer konkurrieren kann. Der einzige unabhängig getestete und in der Literatur berichtete Fall ist ein Selbstversuch von Brady (1970), der einige Jahre später von Carroll nachgetestet wurde (Carroll, 1975).

Diese Eigenschaft des AGs unterscheidet es deutlich von den gängigen Konzeptionen von Gedächtnissystemen in verschiedenen Domänen, die in der Regel als flexibel, trainier- und modifizierbar gelten (vgl. die definitorische Eingrenzung des Begriffes ‚Gedächtnis‘ unter 3.1.3). Populäre Beispiele für leistungsfähige und trainierbare Gedächtnissysteme fanden sich in der Vergangenheit zur Genüge in der reichweitenstarken Fernsehsendung *Wetten*

dass ...?, bei der Teilnehmer nach langem Training in der Lage waren, z.B. Zahlen von mehreren hundert Stellen zu memorieren, Daten zu allen Päpsten aus dem Gedächtnis wiederzugeben, eine Unzahl von Barcode-Mustern wiederzuerkennen oder Mineralwässer am Geschmack zu identifizieren. Wie lächerlich würde eine Wette (nicht nur für Absoluthörer) wirken, in der z.B. 48 verschiedene Töne an ihrer Tonhöhe erkannt werden sollen, und das, obwohl der Teilnehmer sich vielleicht genauso lange und intensiv auf diese Aufgabe vorbereitet hat, wie der Barcode-Spezialist.

Im Gegensatz zur Gedächtniskonzeption, die die Beteiligung höherer kognitiver Systeme voraussetzt, kann das AG sehr gut in Analogie zum Farbsehen gebracht und damit als Fähigkeit des Wahrnehmungsapparates klassifiziert werden. Das Absoluthören zeigt denn auch einige Parallelen zum Erkennen und Benennen von Farben:

Absoluthörer müssen in der Regel ihre Fähigkeit nicht (bewusst) trainieren.

Die Fähigkeit besteht darin, einer Wellenlänge (bzw. Frequenz) eine eindeutige verbale Bezeichnung zuzuordnen.

Das primäre Hörzentrum im Cortex ist – noch auf viel einfachere Weise als der primäre visuelle Cortex – entlang der physikalischen Dimension geordnet.

Einige Konzepte zum Erlernen des AG zielen deshalb auch bewusst auf das Training der Wahrnehmungsfähigkeit für Tonchromata hin und vernachlässigen ein kognitives Gedächtnistraining für Tonhöhen (z.B. Burge, 1986).

Typischerweise werden in Experimenten zum AG die Probanden gebeten, die Tonnamen von einzelnen oder zusammenhängenden Sinus- oder Komplextönen zu nennen oder genannte Tonhöhen zu reproduzieren. Reproduktionsstudien sind aber insgesamt seltener, da die Analyse und Klassifizierung der wiedergegebenen Tonhöhen schwieriger ist als das Feststellen der Korrektheit von verbalen Angaben (vgl. die ähnliche Problematik für *recognition*- und *reproduction*-Paradigma in Experimenten zum Gedächtnis für melodische Inhalte in 3.2 und 3.3). Umfangreiche und bekannte Reproduktionsstudien zum absoluten Gehör sind z.B. Halpern (1989) oder Bergson & Trehub (2002).

Die üblichen abhängigen Variablen, die in Experimenten zum AG erhoben werden, sind der Anteil der korrekten Antworten sowie die Art der Fehler und die Reaktionszeit für die Antworten. Denn einige Konzeptionen des AGs gehen davon aus, dass die Fähigkeit sich auch in der Spontaneität der Antworten manifestiert, da das Erkennen der Tonhöhen für den Absoluthörer mühelos sein muss. In Hinsicht auf den Fehlertyp lassen sich mehrere

Variablen aus den Antworten der Versuchspersonen ableiten, so z.B. die Verteilung der genannten Tonhöhen (in Relation zur korrekten Tonhöhe), die Differenz (modulo 12) zwischen den genannten und den korrekten Tonhöhen, die Richtung der Abweichungen, der Zusammenhang zwischen den Antworten der Probanden etc. Typische unabhängige Variablen in AG-Experimenten sind: die musikalische Vorbildung der Probanden, die Art der Teststimuli (Einzeltöne bis ganze Musikstücke) und die Art der Störbedingung.

So einfach dieses experimentelle Grunddesign sein mag, so schwierig ist es, die Identifizierung und Quantifizierung des absoluten Gehörs eindeutig zu operationalisieren. Diese Schwierigkeit ist vor allem zweifach begründet:

Erstens muss sichergestellt sein, dass die Leistungen in einem Test wirklich und ausschließlich Leistungen sind, die auf die Fähigkeit des Absoluthörens zurückgehen. Denn vor allem, mit dem unter Musikern weit verbreiteten und oft gut trainierten relativen Gehör, also der Fähigkeit, Intervalle oder Mehrklänge relativ zu einer gegebenen Referenztonhöhe zu bestimmen, können in Testsituationen Tonidentifikationsleistungen erbracht werden, die dem absoluten Gehör vergleichbar sind. Denn wie Rakowski und Morawska-Büngeler (1987) angeben, kann eine einzelne Tonhöhe sehr lange (mindestens fünf Minuten, s.a. den folgenden Abschnitt 3.1.2) im Gedächtnis behalten werden, wenn keine Störreize vorhanden sind. Die meisten experimentellen Studien zum AG verwenden deshalb Störreize in irgendeiner Form, z.B. Akkordmodulationen durch viele unterschiedliche und entfernte Tonarten zwischen den einzelnen Versuchsdurchgängen oder sprachliche, kognitive oder gesangliche (Reproduktions-)Aufgaben der Probanden (vgl. z.B. die Kombinations- und Rechenaufgaben von Heyde (1987, S. 263-267)). Als effektivste und sicherste Störvariable bezeichnet Vitouch jedoch die Zeit, die zwischen zwei Versuchsdurchgängen liegt. In einer Studie von Vitouch und Gaugusch (2000) liegen deshalb jeweils 24 Stunden zwischen den Versuchsdurchgängen. Dieses zeitintensive Vorgehen ist jedoch in der Regel für umfangreiche Untersuchungen nicht praktikabel. Die große Bedeutung der Störreize in der Forschungspraxis zum AG zeigt, wie wichtig die Ausschaltung aller Gedächtniskomponenten (zumindest des Kurzzeitgedächtnisses) für die Bestimmung und operationale Definition des AGs ist. Das AG soll gerade keine Fähigkeit sein, die sich in irgendeiner Form auf das Kurzzeitgedächtnis stützt. Ob das AG als eine besondere Erscheinungsform des Langzeitgedächtnisses betrachtet werden kann, wird unten bei der Diskussion der Studien von Levitin und anderen erörtert.

Zweitens muss die Auswertung von Art und Anzahl der Fehler in einem Experiment dergestalt sein, dass zwischen Absoluthörern und Nicht-Absoluthörern unterschieden werden

kann. Denn auch Personen, die zweifelsfrei das AG besitzen, machen in AG-Experimenten Fehler. Diese Fehler gehören jedoch i.d.R. zu einer von zwei Klassen: a) Identifizierung des richtigen Tonchromas, aber der falschen Oktavlage (Oktavfehler) und b) fälschliche Identifizierungen von Tönen, die einen Halbton über oder unter der richtigen Tonhöhe liegen. Aufgrund der Stabilität dieses Fehlermusters werten viele Studien falsche Tonidentifizierungen (oder Produktionen), die zu einer oder zu beiden der genannten Fehlerklassen gehören, nicht als Fehler, sondern als korrekte Antwort (vgl. die Überblickstabelle, die Takeuchi und Hulse (1993) zur Behandlung dieser Fehlerarten geben). Auch mit der Oktav- und Halbtontoleranz, so wird in vielen Studien argumentiert, lassen sich Absolut Hörer immer noch sicher als solche identifizieren.

In den aktuelleren Studien wird jedoch häufig zwischen verschiedenen Typen bzw. Ausprägungen des AGs unterschieden. Eine Hypothese, die schon früh in die Theoriebildung zum AG eingebracht wurde (z.B. Weinert, 1929; Revesz, 1946), geht davon aus, dass das AG kein dichotomes Merkmal ist (vorhanden vs. nicht-vorhanden), sondern dass die Fähigkeit, Tonhöhen zu identifizieren, kontinuierlich ausgeprägt ist. Für diese Hypothese spricht vor allem, dass es bisher nicht gelungen ist, ein allgemeingültiges Merkmal für die Klassifizierung von Probanden in AG-Besitzer und nicht-AG-Besitzer zu etablieren. So zeigen z.B. in der von Bischoff et al. (2000) getesteten Gruppe von Musikern mit und ohne AG (nach a priori Selbsteinschätzung) einige Versuchspersonen ohne AG sehr gute Tonidentifizierungsleistungen, die den Leistungen von ‚echten‘ Absolut Hörern vergleichbar sind. Auch die Ergebnisse von Hurni-Schlegel und Lang (1978) sowie von Vitouch und Gaugusch (2000) verweisen nach der Interpretation der Autoren selbst in Richtung der Hypothese, dass das AG eine kontinuierliche Fähigkeit darstellt.

Andere Forscher verwenden zwar nicht den Begriff des Kontinuums, der zumindest implizit Vorstellungen von Äquidistanz und infiniter Teilung der Messskala nahe legt, plädieren aber für die verwandte Vorstellung eines Stufenmodells des AGs (z.B. Taneda & Taneda, 1993). So können verschiedene (Teil-)Fähigkeiten eingeordnet werden, die oft bei der Untersuchung des AGs beobachtet werden. Dazu gehören z.B. die Identifikation von Einzeltönen auf einem Instrument (oder einer Instrumentengruppe), mit dem die Vpn. die meiste praktische Erfahrung gesammelt hat; auch die Identifikation von Tönen in nur bestimmten Bereichen des musikalischen Spektrums (nur ‚weiße‘ Tasten oder nur Töne in kleiner und eingestrichener Oktave) gehört zu den vermeintlich niedrigeren Stufen dieser Stufenmodelle. Die höheren Stufen wären nach Taneda und Taneda die unbeschränkte Identifikation und Wiedergabe aller möglichen Tonhöhen unter allen möglichen Bedingungen. Wie

Vitouch in seiner Zusammenschau schreibt, ist jedoch nicht entschieden, ob hierarchische Stufenmodelle, nach denen bestimmte Stufen durchlaufen werden müssen, um die nächsthöheren zu erreichen, angemessen sind, oder ob nicht vielmehr mehrdimensionale *trait*-Modelle – wie sie z.B. klassischerweise für die Definition und Messung von Intelligenz verwendet werden – für die modellhafte Erklärung der vielen verschiedenen Formen des AGs, die in der Realität anzutreffen sind, mehr Erklärungswert besitzen.

Da – wie gesagt – die Ergebnisse von experimentellen Studien zum AG in der Regel unter penibler Ausschaltung aller möglichen Komponenten und Subsysteme des Kurzzeitgedächtnisses zustande gekommen sind, sind sie im Rahmen dieser Arbeit in der Mehrzahl von geringer Bedeutung. Von größerem Interesse sind Untersuchungen, die den menschlichen Umgang mit absoluten Tonhöhen unter einer Gedächtnisperspektive betrachten.

Dabei werden im folgenden Teilkapitel zunächst Untersuchungen betrachtet, die die Leistungen des Kurzzeitgedächtnisses für Einzeltöne erörtern. Im anschließenden dritten Teilkapitel sollen solche Untersuchungen zusammengefasst werden, die das Langzeitgedächtnis für absolute Tonhöhen von sehr bekannten Musikstücken untersuchen.

3.1.2 Gedächtnis für Einzeltöne

Wie im vorangegangenen Abschnitt bereits angedeutet, können auch Personen ohne absolutes Gehör eine einzelne Tonhöhe für eine gewisse Zeit absolut identifizieren, wenn keine Störreize vorliegen. Diese Fähigkeit wird generell als Leistung des Kurzzeitgedächtnisses oder des Arbeitsgedächtnisses interpretiert, welche in ihrer jeweiligen Konzeption (Atkinson & Shiffrin, 1968 bzw. Baddeley, 1986) als in ihrer Kapazität begrenzt und durch Interferenzen störbar angenommen werden (vgl. auch Baddeley, 1997).

Es stellen sich die Fragen, welcher Art die Interferenz sein muss, um die Repräsentation einer absoluten Tonhöhe im Gedächtnis zu löschen oder zu stören und welches die Grenzen des akustischen Speichers für Tonhöheninformationen sind. Diesen Fragen widmete sich Diana Deutsch in einer Reihe von Studien in den 70er Jahren (Deutsch, 1970, 1972, 1974, 1975a, 1975b; Deutsch & Feroe, 1975), die zudem noch einige Nachfolgestudien anderer Forscher in den darauf folgenden Jahren inspirierten (z.B. Kallman et al., 1987; Butler & Ward, 1988; Semal & Demany, 1991; Pechmann & Mohr, 1992).

Die folgende Zusammenfassung orientiert sich im Groben an der Darstellung bei Deutsch (1999), die sowohl die eigenen Ergebnisse referiert und zu einer Gedächtnistheorie verbind-

det als auch die späteren Forschungen anderer Wissenschaftler einbezieht. Zusätzlich sind die Ergebnisse einer Reihe von weiteren Studien eingearbeitet, die Deutschs Ergebnisse differenzieren.

In einer Initialstudie von 1970 testet Deutsch das Tonhöhengedächtnis in einem Wiedererkennensparadigma (vgl. die detaillierte Beschreibung des experimentellen Paradigmas in Kap. 3.2). Dabei wird den Versuchspersonen ein Ton einer bestimmten Tonhöhe zu hören gegeben, der im Gedächtnis behalten werden soll. Nach einer Störbedingung in der Retentionsphase, die in diesem Experiment aus Tönen anderer Tonhöhen und gesprochenen Zahlen besteht und die Deutsch Interpolationsbedingung nennt, wird eine Vergleichstonhöhe dargeboten. Die Aufgabe der Versuchspersonen ist es zu beurteilen, ob der erste Ton und der Vergleichston die selbe Tonhöhe besitzen oder nicht.

An der Zeit (dem Retentionsintervall), die zwischen den beiden zu vergleichenden Tönen liegt, ist Deutsch nicht interessiert, denn in vielen Experimenten der 50er und 60er Jahre hatte sich bereits als stabiler Befund ergeben, dass das Gedächtnis für eine einzelne Tonhöhe mit der Zeit schwindet. Deutsch ist in diesem Experiment vielmehr am Einfluss der Störbedingungen interessiert.

Sie findet eine deutliche Interferenz nur mit den Tönen in der Störbedingung, nicht mit dem sprachlichen Material. Dabei steht die Anzahl der in der Störbedingung interpolierten Töne in positiven Zusammenhang mit der Anzahl der Fehler. Deutsch schließt daraus, dass es unterschiedliche Gedächtnissysteme geben muss, die für das Behalten von Zahlen (sprachlichem Material) und Tonhöhen zuständig sind. Das Gedächtnis für Tonhöhen wird nach diesem Ergebnis offensichtlich zu einem kleineren Teil von der Zeitspanne des Retentionsintervalls beeinflusst und zu einem größeren von der Anzahl und der Art der Interferenzen, die während des Behaltens dargeboten werden.

Weitere Hinweise darauf, dass es sich beim Gedächtnis für Tonhöhen tatsächlich um ein eigenes, vom verbalen Gedächtnis weitgehend unabhängiges System handelt, fanden Deutsch und andere Forscher in einer Reihe von Folgestudien, die sich jeweils desselben experimentellen Designs bedienten, jedoch unterschiedliche Töne und Materialien in der Interpolationsphase einsetzten. So findet Deutsch (1974), dass vor allem die Töne der Interpolationsbedingung die Gedächtnisleistung für die Testtonhöhe verschlechtern, die in derselben Oktave wie diese dargeboten werden. Interpolierte Töne in der oberen oder unteren

ren Oktave beeinträchtigen die Leistungen im Versuch deutlich weniger. Die Gedächtnisleistungen hängen also offensichtlich von der Beziehung der Testtöne zu den Störtönen ab. Semal und Demany reproduzierten dieses Experiment zweimal (1991, 1993): Einmal mit obertonreichen Tönen mit komplexen Spektren, denen allerdings der Grundton fehlte, und einmal mit einer Manipulation der Lautstärke der Störtöne. Im Ergebnis fanden sie nur einen minimalen Effekt des Spektrums bzw. der Lautstärke. Offensichtlich wird in dem beanspruchten Gedächtnissystem die Tonhöhe von den übrigen Eigenschaften des Tones abstrahiert und stört dann die Erinnerung an eine Referenztonhöhe, wenn die Störtöne zu dieser in naher Beziehung stehen, und zwar im Sinne der Grundfrequenz, die jedoch nicht einmal physikalisch im Signal präsent sein muss.

Genauso unerheblich scheint der Ort der Schallquelle für die Leistungen in diesem experimentellen Paradigma zu sein zu sein. Wie Deutsch (1978) zeigen und Kallmann et al. (1987) bestätigen konnten, bedeutet es für die Vpn. nur einen minimalen Vorteil, wenn Vergleichs- und Störtöne je einem Ohr dargeboten werden.

Nimmt man zur Erklärung dieser Befunde ein spezielles Kurzzeitgedächtnissystem für (abstrahierte) Tonhöhen an, dann scheint dieses mit dem Grad an musikalischer Vorbildung oder Aktivität stärker ausgeprägt zu sein. Pechmann und Mohr (1992) fanden in einem sehr ähnlichen experimentellen Paradigma eine leichte Beeinträchtigung der Tonhöhen-Wiedererkennensleistungen bei Nicht-Musikern durch verbales und visuelles Material in der Retentionsphase, bei Musikern jedoch keinerlei Beeinträchtigung durch diese Störmaterialien. Eine Interpretation, die Pechmann und Mohr selbst liefern und die auf dem *Working Memory*-Modell von Baddeley fußt (Zusammenfassung z.B. bei Müllensiefen, 2000b), sind zwei getrennte auditorische Subsysteme: eine artikulatorische Schleife, wie sie Baddeley beschreibt, und als ‚Schwestersystem‘ eine tonale Schleife, die aus einem begrenzten tonalen Speicher und einem *rehearsal*-Mechanismus besteht. Diese tonale Schleife wäre demnach bei Musikern stärker ausgeprägt, in dem Sinne, dass der Eingang des tonalen Materials in den Speicher und das Wiederholen des Materials im *rehearsal*-Mechanismus weitgehend automatisiert sind und nur sehr geringe Ressourcen einer zentralen Aufmerksamkeitskomponente verbrauchen. Bei Nicht-Musikern wäre bei der gleichzeitigen Bearbeitung von sprachlichem und tonalem Material mehr Aufmerksamkeit erforderlich, sodass die Behaltensleistungen schlechter werden.

Allerdings konnte Münzer (1998) in einer Variante dieses Experimentes, bei der Musiker und Nicht-Musiker sprachliches Material unter Störung von tonalen Reizen behalten sollten, keine eindeutige Dissoziation eines verbalen und eines tonalen Subsystems finden.

Das bedeutet: In der Interpretation seiner Daten konnte Münzer nicht zwischen einem besser ausgebildeten speziellen tonalen Subsystem bei Musikern oder einer einfach größeren Ressource des auditorischen Gedächtnisses bei Musikern allgemein unterscheiden.

Einzelne Tonhöhen scheinen also für das Behalten vom konkret erklingenden Ton abstrahiert und in einem eigenen Subsystem gespeichert zu werden. Wie stark oder scharf ist jedoch die Abstraktion bzw. Kategorisierung auf der Dimension der Tonhöhe? Diese Frage zielt auf die Bedeutung von Tonhöhenabständen im Gedächtnis, die nicht unbedingt unseren musikalischen Tonabständen entsprechen müssen. Deutsch untersuchte diese Frage mit demselben experimentellen Paradigma (Deutsch, 1972), wobei sie besonderen Wert auf einen kritischen Ton legte, der an zweiter Stelle der interpolierten Töne zu hören war. War dieser zweite interpolierte Ton identisch zum Referenzton, so waren die Leistungen der Versuchspersonen besser. Der kritische Ton wurde offensichtlich als Auffrischung der im Gedächtnis befindlichen Tonhöhe verwendet. Wicht der kritische Ton jedoch um einen geringen Betrag vom Referenzton ab, so waren die Leistungen schlechter. Die meisten Fehler generierte dabei ein kritischer Ton, der um einen $2/3$ -Ton (ca. 133 Cent) vom Referenzton abwich. Dieser Wert erschien als Maximum einer unimodalen Verteilung der Fehler, die bei kleineren und größeren Abständen zum Referenzton abnahmen.

Zur Erklärung dieses Zusammenhangs zwischen Frequenzabständen und Gedächtnisfehlern zieht Deutsch eine Analogie zum Mechanismus der lateralen Hemmung in anderen Bereichen der sensorischen Verarbeitung (z.B. Konturwahrnehmung im visuellen System): Die Dimension der Tonhöhe sei im menschlichen auditorischen System entlang eines logarithmisch skalierten Kontinuums organisiert. Die Aktivierung einer bestimmten Tonhöhe hemme nun gleichzeitig die benachbarten Tonhöhenrepräsentationen. Bei diesem Prozess sind die Bandbreiten von Bedeutung, inner- bzw. außerhalb derer eine Hemmung stattfindet oder nicht stattfindet. Frequenzen, die sehr nahe (z.B. im Abstand kleiner als ein Halbton) beieinander liegen, hemmen gar nicht oder gering, da dieselbe Tonhöhenkategorie aktiviert wird. Die größte Hemmung wird durch eine Frequenz hervorgerufen, die im Abstand von $2/3$ eines Ganztones zur Referenztonhöhe liegt, deutlich größere oder kleinere Störfrequenzen aktivieren andere Tonhöhenkategorien im Gedächtnis, die entweder parallel zum Referenzton behalten werden oder als von diesem verschieden und damit für die Aufgabe irrelevant beurteilt werden können. Diese kritische Bandbreite für Störungen der Gedächtnisrepräsentation von Tonhöhen entspricht sehr gut der Bandbreite für die laterale Hemmung im auditorischen System, wie sie beispielsweise Klinke et al. (1970) angeben.

Ebenso wie bei der neuronalen lateralen Hemmung kumulieren die Fehlerraten, wenn zwei kritische Töne ober- und unterhalb des Referenztones dargeboten werden, wie Deutsch später (1973) zeigen konnte.

Einen weiteren starken Hinweis auf die Validität dieses inhibitorischen Gedächtnismechanismus für Tonhöhen fanden Deutsch und Feroe (1975) durch einen Test auf Desinhibition, der wiederum dasselbe experimentelle Paradigma verwendete. Der zweite kritische Ton wurde im Frequenzabstand relativ zum ersten kritischen variiert. Als folgerichtiges Ergebnis fanden Deutsch und Feroe, dass die Gedächtnisleistung für die Testtöne dann am schlechtesten war, wenn beide kritischen Töne im Abstand eines 2/3-Tones zum Referenzton standen. War der zweite kritische Ton jedoch einen 2/3-Ton vom ersten kritischen Ton entfernt, dann war die Gedächtnisleistung für den Referenzton am besten. Die Repräsentationen der kritischen Töne hemmten sich also gegenseitig und behinderten dadurch die Erinnerung an den Referenzton weniger. Standen erster und zweiter kritischer Ton im Abstand von 2/3 eines Ganztones, so waren die Erinnerungsleistung sogar besser als in der Bedingung mit nur einem kritischen Ton. Deutsch und Feroe folgern deshalb, dass durch den zweiten kritischen Ton die Inhibition des ersten kritischen Tones verhindert wird.

Bis hierhin wurden nur die (richtigen) Antworten der Versuchspersonen in diesem Paradigma betrachtet, wenn der Referenzton zu Beginn und der Testton am Ende identisch waren. In der Terminologie der Signal Detection Theory (vgl. ausführlichere Erläuterungen in Kap. 3.2) waren die *hits* und die *misses* von Bedeutung. Aber auch die richtigen und falschen Antworten bei nicht-identischen Testtönen geben Aufschluss über die Funktionsweise des Tonhöhengedächtnisses. Enthält die Sequenz der interpolierten Töne einen Ton, der identisch zum zweiten Testton ist, dann wird dieser Testton häufig fälschlicherweise als gleich dem Referenzton bezeichnet (Deutsch 1970, 1972; Butler & Ward, 1988). Dieser Effekt war umso stärker, je früher der zum zweiten Testton identische Ton in der Serie der interpolierten Töne zu hören war. Nach Deutschs Interpretation ist das Gedächtnis für Tonhöhen also in zwei Dimensionen organisiert, einer tonotopen geordneten Dimension entlang des Frequenzkontinuums und einer zeitlichen Dimension, die jedoch wohl schlechter in Sekunden zu messen ist als in der Anzahl relevanter Störreize in einem Zeitintervall. Die modellhafte Interpretation, die Deutsch gibt, ist die, dass eine bestimmte Tonhöhe für ein bestimmtes Zeitintervall im Gedächtnis repräsentiert ist. Wird diese Repräsentation innerhalb dieses Intervalls durch erneutes Hören wieder aufgefrischt, so bleibt sie länger und stabiler im Gedächtnis und die Fehler bei der Identifizierung des zweiten

Testtones nehmen ab. Diese Interpretation lässt sich durch die Ergebnisse der Studien von Deutsch aus den Jahren 1970, 1972 und 1975 stützen.

Der Frage, wie die Dimension der Tonhöhe genau im Tonhöhengedächtnis organisiert ist, ging Deutsch in ihrer Studie von 1973 nach. Sie versetzte die kritischen Töne in die obere und untere Oktave relativ zu den Testtönen. Es zeigten sich die gleichen Effekte der Inhibition wie bei der Verwendung von kritischen Tönen in der selben Oktave, jedoch waren diese deutlich geringer. Die kritischen Töne in der tiefen Oktave hatten dabei die geringsten Interferenzeffekte und die in der oberen Oktave produzierten etwas stärkere Effekte. Die Tonhöhe wird also sowohl als bestimmte absolute Tonhöhe in einer Oktave als auch als zugehörig zu einem Tonchroma erinnert.

Abschließend zu dieser kurzen Zusammenstellung der Ergebnisse aus dem ‚Deutsch-Paradigma‘ soll noch ein Befund referiert werden, der in scheinbarem Widerspruch zu dem zuvor dargestellten Ergebnis aus der 1974er-Studie steht und gleichzeitig Hinweise darauf gibt, dass das hier postulierte Tonhöhengedächtnis wohl durch den Erfahrungshorizont der Versuchspersonen in der sprachlichen und/oder musikalischen Domäne geprägt ist. In den übrigen Experimenten der Veröffentlichung von 1974 und in Experimenten, deren Ergebnisse Deutsch 1978 publizierte, platzierte Deutsch die sechs interpolierten Töne in insgesamt drei Oktaven (obere, untere und gleiche wie Testtöne) und variierte gleichzeitig die Intervallbeziehungen zwischen den interpolierten Tönen (auf- oder absteigend vs. Zufallsintervalle). Es zeigte sich, dass dann die Gedächtnisleistungen am schlechtesten waren, wenn die Intervalle zwischen den interpolierten Tönen am größten waren. Die schlechtesten Leistungen wurden bei einem durchschnittlichen Intervall von zwei Oktaven und einer zufälligen Intervallfolge erzielt, die besten Leistungen bei einer geordneten Intervallfolge, die sich innerhalb der oberen oder unteren Oktave relativ zu den Testtönen bewegt. Ähnliche Ergebnisse erhielten auch Olson und Hansen (1977) und Deutsch und Boulanger (1984). Diese Resultate lassen darauf schließen, dass die Versuchspersonen die interpolierten Töne als melodisches Gebilde wahrnehmen und in Beziehung zum Referenzton setzen, auch wenn die Versuchsinstruktion ausdrücklich fordert, die interpolierten Töne zu ignorieren. Ein sinnvolles melodisches Gebilde in einer anderen Oktave lässt sich gut von den Testtönen unterscheiden; ist es in der selben Oktave dargeboten, so treten offenbar leichter Interferenzen auf. Jedoch können in der Regel sowohl der Testton als auch die interpolierten Töne als eigene Sinneinheiten (es drängt sich hier die Modellvorstellung des

chunk als Gedächtniseinheit auf (Miller, 1956)) im Tonhöhengedächtnis repräsentiert und je nach Störbedingung mehr oder weniger gut unterschieden werden. Können die interpolierten Töne jedoch nicht als eine sinnvolle melodische Einheit verarbeitet werden, wie es bei der Streuung über mehrere Oktaven der Fall sein dürfte, dann ist entweder die Verwechslung eines interpolierten Tones mit dem Testton wahrscheinlicher oder die Repräsentation des Testtones wird aus Kapazitätsgründen aus dem Kurzzeitspeicher gelöscht. Denn sechs einzelne Tonhöhen dürften nach Deutschs Annahme mehr Gedächtnisressourcen verbrauchen als eine Tonfolge mit einem einfachen Aufbau. Dieser hier nur implizite Gedanke weist in Richtung der von Deutsch und Feroe später (1981) entwickelten Kodierung von Tonfolgen im Gedächtnis mit Hilfe von (einfachen) rekursiven Regeln.

Sehr deutlich wird an diesem Ergebnis – welches die Behaltensleistung für eine Tonhöhe von der musikalischen Sinnfälligkeit des Kontextes der interpolierten Störtöne abhängig macht –, dass ein ‚reines‘ Gedächtnis für Einzeltöne nur unter ganz bestimmten experimentellen Bedingungen beobachtbar ist; nämlich dann, wenn alle anderen musikalisierbaren Reize ausgeschlossen sind. Im anderen Fall scheint die kognitive Verarbeitung eine gegebene Tonhöhe unwillkürlich in Bezug zu anderen Tonhöhen zu setzen und diese anderen Tonhöhen als *cues* beim Behalten der Testtonhöhe zu nutzen. Dann kommen jedoch eine Reihe von weiteren Variablen ins Spiel, die das Gedächtnis für Tonhöheninformationen bedingen und miteinander interagieren.

Dieser Schluss scheint für Deutsch zu bedeuten, dass spezifische Gedächtnissubsysteme auch für die übrigen reiznahen Dimensionen der Musikwahrnehmung postuliert und ihre Spezifika und Kapazitäten beschrieben werden können. Um etwas über das Gedächtnis für musikalische Inhalte zu erfahren, sind dann schließlich die Interaktionen zwischen den reiznahen Gedächtnissubsystemen empirisch zu ermitteln, wie sie bei der Erinnerung von echter Musik oder echten Melodien zwangsläufig zustande kommen. Die Zusammenfassung der Literatur geschieht bei Deutsch (1999) auch nach diesem Schema: Nach einem recht ausführlichen Teil über die Gedächtnisexperimente zu Einzeltonhöhen schließt sie nachfolgend in je einem Abschnitt Erkenntnisse zum Gedächtnis für Timbreinformationen, für Tonhöhenbeziehungen, für Tondauern und schließlich zu den Interaktionen der verschiedenen Gedächtnissubsysteme im Erinnerungsprozess an.

Es lässt sich jedoch durchaus hinterfragen, ob diese Strategie, die Gedächtnisleistungen für musikalische *low-level*-Dimensionen in relativ künstlichen experimentellen Paradigmen zu ermitteln und diese dann zu kombinieren, zu sinnvollen und validen Ergebnissen hinsicht-

lich der Gedächtnisleistung für echte Melodien oder Musik führen kann. Denn es ist keineswegs bedenkenlos vorauszusetzen, dass die funktionalen Beziehungen zwischen den Eigenschaften der *low-level*-Dimensionen und den Gedächtnisleistungen auch bei Kombination mehrerer Dimensionen gelten und sich die Effekte in einer nachvollziehbaren Weise modellieren lassen (idealerweise linear). Bezeichnend ist beispielsweise, dass keines der komplexeren Modelle zum Musik- und Melodiegedächtnis (z.B. Boltz & Jones, 1986; Dowling, 1986; Monahan et al., 1987; Crowder, 1993; Berz, 1995) auf die zahlreichen Ergebnisse aus den Studien substantiell zurückgreift, die in diesem Abschnitt dargestellt wurden.

Zusätzlich zu dieser Zurückhaltung anderer Autoren, Deutschs Ergebnisse für die Theorie- und Hypothesenbildung zum Melodiegedächtnis zu verwenden, ist es die geringe Plausibilität des nur implizit von Deutsch angedeuteten atomistischen Modells – von dem Behalten von Einzeltönen zum Gedächtnis für echte Melodien zu gelangen –, die eine Verwendung der hier referierten Ergebnisse in Modellen zum Gedächtnis für komplexere melodische Materialien wenig sinnvoll erscheinen lässt.

Im folgenden Abschnitt sollen einige Studien referiert werden, die sich mit dem Langzeitgedächtnis für die absolute Tonhöhe echter Musikstücke beschäftigen und somit den Bogen zum eingangs behandelten absoluten Gehör schlagen.

3.1.3 Langzeitgedächtnis für absolute Tonhöhen echter Musik

Wie unter 3.1.1 dargestellt, ist das absolute Gehör eine seltene Fähigkeit und wie aus den Erläuterungen zu den Studien unter 3.1.2 deutlich wurde, ist das Kurzzeitgedächtnis für einzelne Tonhöhen sehr leicht störfähig. Aus diesen beiden Tatsachen ließe sich für die überwiegende Mehrheit der Bevölkerung schließen, dass die Zuordnung von bestimmten absoluten Tonhöhen zu bestimmten musikalischen Gebilden (z.B. ganze Musikstücke) zufallsverteilt ist oder einem nicht-musikalischen Prinzip folgt (z.B. einer individuell präferierten konstanten Tonhöhe). Es sind jedoch in den letzten 15 Jahren eine Reihe von Studien publiziert worden, die zeigen, dass auch ‚Normalhörer‘ zumindest innerhalb gewisser Grenzen die absolute Tonhöhe von gut bekannten Musikstücken über lange Zeit relativ stabil im Gedächtnis behalten können. Die Ergebnisse dieser Studien sollen in diesem Abschnitt kurz referiert werden, um abschließend den Unterschied dieser Gedächtnisleistungen zu den Fähigkeiten von Absoluthörern darzustellen.

Als umfangreiche Initialstudie kann in diesem Zusammenhang die Veröffentlichung von Halpern (1989) gesehen werden, die genau der eingangs erwähnten Frage nachging, wie gut Nicht-Absoluthörer die absolute Tonhöhe von bekannten Liedern reproduzieren und wiedererkennen können. Der große Vorteil dieser Studie ist die Tatsache, dass Halpern in vier Experimenten vier verschiedene experimentelle Methoden verwendet, um die Repräsentationen von absoluten Tonhöhen im Gedächtnis der Versuchspersonen zu untersuchen: Neben der wiederholten Reproduktion von Tonhöhen mit der eigenen Stimme bzw. auf einem Keyboard, verwendet sie auch eine *rating*-Prozedur, bei der die Versuchspersonen angeben müssen, wie gut einzelne Töne als Starttöne für gewisse Lieder ‚passen‘ und schließlich noch eine Prozedur, bei der die Versuchspersonen Transpositionsgrenzen einzelner Lieder festlegen müssen. Obwohl im Einzelnen das methodische Vorgehen von Halpern sicher diskutierbar ist – z.B. werden in Experiment 3 und 4 die Daten der Versuchspersonen relativ früh gemittelt, sodass individuelle Bewertungsmuster nicht betrachtet werden können –, liefern die vier Experimente in der Hauptsache doch konvergente Ergebnisse: Die Repräsentation bekannter Lieder beruht bei musikalisch nicht vorgebildeten Personen nicht ausschließlich auf relativen Prinzipien, sondern es wird auch absolute Tonhöheninformation ‚mitgespeichert‘. So war die Streuung der Starttonhöhen ein und desselben Liedes in wiederholten und zeitlichen entfernten Versuchen deutlich geringer als die Streuung der Starttonhöhen verschiedener Lieder: Die mittlere Standardabweichung für die Starttöne derselben Lieder betrug 1,28 Halbtöne im vokalen Reproduktionsversuch und 2,36 Halbtöne bei der Reproduktion auf einem Keyboard. Dies ist zwar deutlich mehr als die Variabilität, die Absoluthörer in vergleichbaren Reproduktionsaufgaben zeigen (Abweichung von maximal einem Halbton), aber gleichzeitig deutlich geringer als die Variabilität der Starttöne unterschiedlicher Lieder (2,52 Halbtöne bei vokaler und 4,03 Halbtöne bei manueller Reproduktion). Damit lässt sich ausschließen, dass sich unter Halperns Versuchspersonen nur sogenannte ‚Brummer‘ befanden, deren Reproduktionsfähigkeit zumindest stimmlich eingeschränkt ist. Auch die Ergebnisse der Transpositionsaufgabe (Exp. 4) weisen darauf hin, dass Nicht-Absoluthörer für bekannte Lieder zumindest präferierte Tonregionen haben. Sie weist selbst daraufhin, dass diese präferierten Tonregionen nicht mit der Fähigkeit eines Absoluthörers vergleichbar sind: „[...] ordinary people store absolute pitch to a greater extent than is commonly believed, this should not be taken to mean that everyone has a mild case of what is traditionally called AP.“ (Halpern, 1989, S. 580).

Wodurch sich diese präferierten absoluten Tonregionen ergeben, versucht Halpern jedoch nicht aufzuklären. Denkbar wären neben dem häufigen Hören oder Singen der Lieder auf

einer bestimmten Tonhöhe z.B. auch der Formantbereich der vorherrschenden Vokale der Liedtexte (das Lied „Row“ war z.B. in Halperns Experiment 4 dasjenige mit der tiefsten Tonregion).

Während Halpern als Versuchsmaterial Volks- und Kinderlieder verwendete, kommt Levitin (1994) zu einem sehr ähnlichen Ergebnis mit Popstücken, die die Probanden selbst wählen können und die ihnen sehr gut bekannt sind. Er ließ seine Versuchspersonen beliebige Ausschnitte aus je zwei Liedern singen, die nach Möglichkeit nur in einer einzigen Aufnahme existieren und daher den Versuchspersonen nur mit einer bestimmten absoluten Tonhöhe bekannt sein konnten, und verglich die gesungenen Tonhöhen mit den Originaltonhöhen der Gesangsstimme auf der Aufnahme. Nach der Reduktion auf eine Oktave (modulo 12) fand Levitin erstaunlich geringe Abweichungen von der Originaltonhöhe: In Versuch 1 sangen 67% der Versuchspersonen eine Tonhöhe, die innerhalb von \pm einem Ganzton um die Originaltonhöhe herum lag (26% trafen die Originaltonhöhe sogar exakt) und in Versuch 2 waren 60% innerhalb eines Ganztones (23% treffen genau). Levitin kann auch stabile individuelle Unterschiede hinsichtlich dieser Reproduktionsfähigkeit feststellen. Ausgehend von einer dichotomen Einteilung der Versuchsergebnisse in *hits* (Tonhöhe exakt reproduziert) und *misses* (Tonhöhe nicht exakt reproduziert), zeigt sich, dass 72% der Versuchspersonen das gleiche Ergebnis in beiden Versuchen erreichen. Dabei sind es nur 12%, die in beiden Versuchen *hits* erreichen und 60%, die bei beiden Malen *misses* landen. Nimmt man als *hit* die Grenzen von \pm einem Ganzton, so waren 44% der Versuchspersonen in der Lage *hits* in beiden Versuchen zu produzieren. Dies lässt Levitin zu der Schlussfolgerung kommen, dass ein weit größerer Teil der Bevölkerung ein absolutes Gedächtnis für musikalischen Tonhöhen („absolute memory for musical pitch“) besitzt als bislang angenommen. Zwar scheint sich Levitin zu scheuen, diesen Menschen direkt das absolute Gehör zuzusprechen – denn beispielsweise war die Benennung und Reproduktion anhand von beliebigen Tönen und Tonnamen nicht nachgewiesen worden –, aber er suggeriert, dass viele Leute ein latentes absolutes Gehör besitzen, aber die erinnerten absoluten Tonhöhen nicht benennen können. Für ihn ist das absolute Gehör somit eine Gedächtnisfähigkeit, die mit verbalem Wissen zusammengeht:

It might be the case that many people possess pitch memory, but never acquired pitch labeling, possibly because they lacked musical training or exposure during a critical period. (Levitin, 1994, S. 421)

Von dieser Hypothese gehen auch einige Trainingsprogramme zum absoluten Gehör (z.B. Burge, 1981) aus, bei denen lediglich die verbale Verknüpfung zu den bereits vorhandenen Gedächtnisinhalten erlernt werden muss.

Die Idee hinter Levitins Erklärungsmodell zum absoluten Gehör bzw. zum absoluten Tongedächtnis ist die einer Prägung durch sehr häufige Wiederholung in der Umgebung. Konsequenterweise wurde nur das Gedächtnis für Töne aus Songs getestet, die den Versuchspersonen sehr gut bekannt waren, d.h. die sie unzählige Male (auf derselben absoluten Tonhöhe) gehört hatten. Ein Experiment, welches Levitin 1997 veröffentlichte, geht der Bedeutung der Umweltprägung in Hinsicht auf die Erinnerung für die Frequenz einzelner Tonhöhen nach. Da bei der fixierten Stimmung des größten Teils der westlichen Musik auf den Kammerton ($a' = 440\text{Hz}$) alle musikalischen Tonhöhen einer festen Frequenzzahl zugeordnet sind, sollten bei der Reproduktion von einzelnen Tönen aus dem Gedächtnis viel häufiger die Töne mit den festen Frequenzen in Relation zum Kammerton produziert werden als Töne zwischen diesen Frequenzen. Tatsächlich findet Levitin zumindest eine teilweise Bestätigung dieser Hypothese bei der Reproduktion mittels eines Tongenerators (Levitin, 1997). Wie Vitouch (2004) berichtet, konnte dieser Befund in einer unabhängigen Replikation durch Ian Cross nicht bestätigt werden.

Es scheint jedoch Vorsicht angebracht, aus Levitins Ergebnissen von 1994 allzu weitreichende Schlussfolgerungen hinsichtlich der Verbreitung des AGs in der Bevölkerung zu ziehen.

Erstens wurden von jeder Vpn. nur zwei Songs aus dem Gedächtnis wiedergegeben. Die zwei Songs gehörten zu den musikalischen Objekten, die der jeweiligen Vpn. am allerbesten bekannt waren. Der Schluss, dass die meisten Menschen sich generell auch die absolute Tonhöhe von Musikstücken merken, ist also nicht aus diesen Daten zu ziehen. Plausibler scheint die Annahme, dass nur bei vielmaliger aktiver Wiederholung die absolute Tonhöhe eines Musikstückes im Gedächtnis gespeichert wird und nur dann, wenn relative Strukturen, wie z.B. die Gesangsmelodie oder melodisch charakteristische Instrumentalteile, schon gespeichert sind. Hierin besteht in jedem Fall ein Unterschied zur Fähigkeit des absoluten Gehörs, wie sie oben definiert ist.

Zweitens besteht in Levitins Interpretation seiner Daten eine gewisse Unklarheit hinsichtlich der Bedeutung des verbalen Benennens („verbal labeling“) von Tonhöhen. Es ist durchaus möglich, dass gerade das Verknüpfen der Wahrnehmung von absoluten Tonfrequenzen mit verbalen Codes das absolute Gehör und Tongedächtnis definiert oder zumin-

dest ein integraler Bestandteil dieser Fähigkeiten ist. Denn Absolut Hörer speichern aller Wahrscheinlichkeit nach nicht den sensorischen Eindruck, sondern sind in der Lage, sensorische Eindrücke absoluter Tonhöhen verbal zu rekodieren und so auch für lange Zeit im Gedächtnis zu behalten. Tatsächlich verwandeln sich durch die verbale Rekodierung die absoluten Wahrnehmungen von Tonhöhe in Wissensobjekte, wie sie traditioneller Weise Bestandteile des Gedächtnisses sind. So definiert beispielsweise Kluwe (1996, S. 196) Gedächtnis wie folgt:

Der Begriff G.[edächtnis] bezieht sich auf verschiedene Formen des Erwerbs, der Speicherung und der Nutzung von Wissen. Mit dem Begriff Wissen (*knowledge*) werden Kenntnisse über die Realität (*declarative knowledge*), kognitive Operationen und Prozesse (*procedural knowledge*) sowie perzeptuell-motorische und kognitive Fertigkeiten (*skills*) bezeichnet.

Nach der gängigen Definition können Absolut Hörer deklaratives Wissen im Bereich der absoluten Tonhöhen erwerben und im Gedächtnis behalten. Personen, die ohne deklaratives Wissen in diesem Bereich zumindest unter gewissen Umständen in der Lage sind, absolute Tonhöhe richtig zu reproduzieren (wie in Halperns und Levitins Experimenten) würde man wahrscheinlich eher eine kognitive Fertigkeit zusprechen. Da den Versuchspersonen ohne absolutes Gehör diese Fähigkeit in der Regel nicht bewusst ist und sie die Fähigkeit nicht von einem konkreten musikalischen Objekt abstrahieren können (zumindest fehlt hierfür bisher der Nachweis), könnte man diese Fähigkeiten im Bereich des sogenannten impliziten Gedächtnisses (vgl. Schacter, 1987) theoretisch einordnen. Absolut Hörer besitzen dagegen ein explizites Gedächtnis für Tonfrequenzen. Diese Unterscheidung ließe sich auch mit Heydes (1987) Interpretation vereinbaren, dass es sich beim absoluten Hören und verwandten Fähigkeiten um Leistungen des Langzeitgedächtnisses handelt.

Wie auch immer die genaue Einordnung innerhalb der unterschiedlichen Gedächtnistheorien bzw. -systeme vorgenommen werden mag, so erhärten die Ergebnisse von Levitins Experiment (1994) in jedem Fall, dass viele Personen gut bekannte Musikstücke mit einer absoluten Tonhöheninformation im Gedächtnis speichern.

Zumindest teilweise konnte auch Hahn die Ergebnisse von Levitin und Halpern in einigen aktuellen Studien bestätigen (Hahn, 2002a). Hahn testete Chorsänger (ohne absolutes Gehör) auf ihre Fähigkeit, die absoluten Tonhöhen der Anfangstöne von ihnen gut bekannten und geübten Chorstücken richtig zu reproduzieren. Zumindest für das erste der getesteten Stücke gaben die Versuchspersonen Tonhöhen wieder, die in der zentralen Tendenz nahe an der Originaltonhöhe lagen (Modus: -1 Halbton) und signifikant von einer Zufallsverteilung abwichen, insgesamt aber zu tief waren. In einer Folgestudie geht Hahn (2002b)

u.a. der Frage nach, wie diese systematische Abweichung nach unten zustande kommt und berücksichtigt deswegen den Faktor ‚Einsingen‘ in einem vierfaktoriellen Design. Doch auch trotz des Einsingens geben die Versuchspersonen systematisch zu tiefe Töne wieder. Allerdings zeigten sich signifikante Korrelationen zwischen der Treffergenauigkeit und den durch die Versuchspersonen subjektiv eingeschätzten Variablen Vertrautheit mit der Melodie und Lernintensität. Dies deutet – wie die Ergebnisse von Levitin für individuell sehr bekannte und beliebte Stücke – darauf hin, dass die vielfache und aktive Reproduktion einer absoluten Tonhöhe eine entscheidende Rolle für die Fähigkeit spielt, die absolute Tonhöhe von Musikstücken annähernd richtig reproduzieren zu können. Dennoch weichen Hahns Ergebnisse deutlich von den enormen Trefferquoten ab, von denen Levitin berichtet.

Auch untersuchte Hahn (Hahn, 2002b) die Bedeutung der aktiven Wiederholung für die Genauigkeit von Wiedergaben absoluter Tonhöhen und verglichen sie mit personenbezogenen Variablen wie musikalischer Vorbildung. Als Testmaterial verwendete Hahn kurze einstimmige und den Versuchspersonen im vorhinein unbekannte tonale und modale Melodien. Es zeigte sich, dass sowohl die Wiederholung bzw. die Dauer des Trainings der Melodien einen Einfluss auf die Akkuratheit der reproduzierten absoluten Tonhöhen hatte, als auch die musikalischen Vorkenntnisse der Versuchspersonen. Eine Entscheidung darüber, ob eher externe Variablen, wie Wiederholung und Trainingsaufwand, oder personenbezogene Variablen, wie Vorbildung und Anlage, von primärer Bedeutung für die Leistungen hinsichtlich der absoluten Tonhöhen verantwortlich sind, wollen die Autoren nicht treffen.

3.1.4 Empirische Ergebnisse der Forschungen zum absoluten Gehör und zum Gedächtnis für Einzeltöne und absolute Tonhöhen ganzer Musikstücke (Tabelle)

Ergebnis	Kommentar	Beleg
Absolutes Gehör		
Prävalenz in Bevölkerung westlicher Kulturen 1:10.000 bis 1:1.000.		Bachem (1955); Profita & Bidder (1988); Takeuchi & Hulse (1993)
Genetische Disposition für Ausbildung des AGs ist bisher nicht zweifelsfrei nachgewiesen, wird aber für wahrscheinlich gehalten.		Zatorre (2003)
Frühe aktive Beschäftigung mit Musik	Als ‚Prägungstheorie‘	Ward & Burns (1982);

ist für Ausbildung von absolutem Gehör sehr förderlich.	bekannt.	Heyde (1987); Vitouch (2004)
Erlernen des AGs im Erwachsenenalter ist fast unmöglich.		Brady (1970); Carroll (1975); Vitouch (2004)
Auch Absolut Hörer machen bei der Identifizierung von Tonhöhen zuweilen Fehler, die aber bestimmten Mustern gehorchen.		Takeuchi & Hulse (1993)
Gedächtnis für Einzeltöne		
Die Erinnerung an eine Tonhöhe kann durch andere Tonhöhen stärker gestört werden als durch sprachliches Material.	Es existiert ein eigenes Gedächtnissystem für Tonhöhen (Deutschs Interpretation).	Deutsch (1970)
Je näher die Frequenzen der Störtöne an der des zu behaltenden Tons liegt, desto stärker ist die Störung.		Deutsch (1974); Semal & Demany (1991); Semal & Demany (1993)
Störeffekte sind weitgehend unabhängig von Spektrum, Lautstärke des Tones sowie Ort der Schallquelle.		Deutsch (1978); Kallman et al. (1987); Semal & Demany (1991); Semal & Demany (1993)
Störeffekte sind dann am größten, wenn der Störton ca. einen 2/3-Ton von der zu behaltenden Tonhöhe entfernt ist.	Effekt wird durch das neuronale Prinzip der lateralen Hemmung generiert (Deutschs Interpretation).	Deutsch (1972); Deutsch (1973); Deutsch & Feroe (1975)
Störeffekte sind auch bei oktavversetzten Störönen mit Abstand +/- eine Oktave +/- 2/3-Ton zu beobachten. Störeffekte sind hier aber geringer.	Tonhöhe wird sowohl nach Tonchroma sowie nach absoluter Tonhöhe erinnert (Deutschs Interpretation).	Deutsch (1973)
Die Gedächtnisrepräsentation für eine Tonhöhe verliert sich mit zunehmender Anzahl von danach dargebotenen Störönen, kann aber durch erneutes Hören des zu behaltenden Tones aufgefrischt werden.	Gedächtnis für Tonhöhen ist in mindestens zwei Dimensionen organisiert: einer tonotopen Dimension und einer zeitlichen (Deutsch Interpretation).	Deutsch (1970); Deutsch (1972); Butler & Ward (1988)
Gedächtnisleistungen für zu behaltende Tonhöhe ist dann am schlechtesten, wenn Intervalle zwischen Störönen besonders groß sind.	Menschen neigen dazu, die zu behaltende Tonhöhe und die Störtöne als melodisches Gebilde wahrzunehmen und zu erinnern.	Deutsch (1974); Olsen & Hansen (1977); Deutsch (1978); Deutsch & Boulanger (1984)
Unabhängigkeit von tonalem und sprachlichem Gedächtnissystem sind bei Musikern deutlicher ausgeprägt.		Pechmann & Mohr (1992)
Langzeitgedächtnis für absolute Tonhöhen echter Musikstücke		
Streuung von wiedergegebenen Starttonhöhen des selben Liedes ist deutlich geringer als Streuung von	Auch Nicht-Absolut Hörer speichern bis zu einem gewissen Grad absolute	Halpern (1989)

Starttonhöhen verschiedener Lieder.	Tonhöheninformation von Musikstücken.	
Viele Menschen reproduzieren absolute Tonhöhen von ihnen gut bekannten Musikstücken (die sie von unveränderlichen Tonaufnahmen kennen) ziemlich exakt (ca. 60% reproduzieren innerhalb +/- 1 Ganzton).	Gilt wahrscheinlich nur für absolute Lieblingslieder, die sehr oft aktiv wiederholt wurden (eigene Interpretation).	Levitin (1994); Hahn (2002a)

3.2 Gedächtnis für kurze Tonfolgen

3.2.1 Bedeutung und untersuchungstheoretischer Ansatz

Kein Aspekt des musikalischen Gedächtnisses ist in der Vergangenheit mehr beforscht worden als das Behalten von kurzen Tonfolgen. Die ersten experimentell-psychologischen Studien zum Melodiegedächtnis mit dieser Art von Stimuli wurden im ersten Drittel des 20. Jahrhunderts durchgeführt (Kochmann 1923; Ortmann 1933). Und für das Konzept des musikalischen Gedächtnisses an sich blieben Studien, wie die von Dowling und Fujitani (1971), Dowling (1978), Idson und Massaro (1978) oder Deutsch (1980) für lange Zeit von großer Bedeutung.

Vor diesem Hintergrund ist es wenig verwunderlich, dass seit 1970 vor allem in den einschlägigen angloamerikanischen psychologischen und musikpsychologischen Fachzeitschriften mehrere Dutzend Studien veröffentlicht worden sind, die alle als experimentellen Stimulus eine oder mehrere Tonfolgen von etwa drei bis zwölf Tönen verwenden.

Diese ‚Beliebtheit‘ von kurzen Tonfolgen als Stimulusmaterial lässt sich zum großen Teil durch den experimentellen Ansatz der meisten Studien erklären, die kurze Tonfolgen als Stimuli wählen. Dieser Ansatz bestand in der Regel darin, Faktoren, Komponenten oder Dimensionen zu identifizieren, die für die Erinnerbarkeit von melodischen Gebilden bestimmend sind (vgl. z.B. Dowlings einflußreichen Artikel „Scale and contour. Two components of a theory of memory for melodies“ von 1978). Die zentrale Frage dieser Studien lässt sich kurz formulieren als: „Was bleibt von einer Melodie oder einer melodischen Phrase im Gedächtnis?“ Um zu einer Aussage über die Bedeutung von bestimmten musikalischen Dimensionen für die Behaltensleistung zu kommen, wird das Stimulusmaterial entlang der möglichen Ausprägungen von einer oder wenigen Dimensionen systematisch variiert, wobei andere musikalische Dimensionen im Versuchsmaterial möglichst eliminiert oder konstant gehalten werden. Mit zwei- oder mehrfaktoriellen Varianzanalysen wird dann im Regelfall nach signifikanten Einflüssen der untersuchten Dimensionen bzw. Variablen auf die Behaltensleistung gesucht. Die zumeist implizite Prämisse der meisten Studien besteht jedoch darin, die für kurze Tonfolgen ermittelten einflußreichen Dimensionen auch als bedeutende Größen für die Erinnerung an Melodien anzunehmen, wie sie in ‚echter‘ Musik in realen Hörsituationen vorkommen. Wie unten im Abschnitt *Ergebnisse* referiert wird, sind dabei als mögliche einflußreiche Dimensionen unter anderem die Bewegungsrichtung, die Tonhöhenkontur, die relativen Intervallgrößen, die exakten Intervallgrößen, die Tonskala und die Tonalität sowie der Rhythmus ausgemacht worden.

Weder dieser Ansatz insgesamt noch die Eignung von kurzen Tonfolgen als Versuchsmaterial innerhalb dieses Ansatzes werden in der experimentellen Literatur selbst oder den zusammenfassenden Texten zu diesem Bereich (z.B. Krumhansl, 1990; Crowder, 1993; Berz, 1995; Deutsch, 1999; Vazan, 2000) unter einem methodischen Gesichtspunkt diskutiert. Eine Ausnahme ist die sehr kurze und eher oberflächliche Rechtfertigung, die Handel (1989, S. 377) für die Experimente mit kurzen Tonfolgen in Hinsicht auf die Verarbeitung echter Melodien findet.

Dennoch lassen sich bei genauerer Betrachtung der Methodik vieler experimenteller Studien einige Eigenschaften von kurzen Tonfolgen ausmachen, die sie für die Verwendung innerhalb des skizzierten Ansatzes qualifizieren. Um dem Leser einen Eindruck von Art und Eigenschaften der diskutierten Versuchsmaterialien zu vermitteln, sind einige Beispiele in Notenschrift hinzugefügt.

3.2.2 Eigenschaften von kurzen Tonfolgen als experimentelle Stimuli

Kurze Tonfolgen weisen zunächst einige Eigenschaften auf, die sie als experimentelle Stimuli von den im vorangegangenen Kapitel besprochenen Einzeltönen unterscheiden:

a) Beziehungen der Töne untereinander: Im Gegensatz zum Gedächtnis für Einzeltöne wird generell angenommen, dass bei der Abfolge von mehreren Tönen die Beziehungen der Töne untereinander für die Behaltensleistung eine Rolle spielen. Auf die Tatsache, dass Einzeltöne und Tonfolgen durch verschiedenartige Prozesse verarbeitet werden, weisen beispielsweise Idson und Massaro (1976) deutlich hin.

Es scheint somit möglich, als unabhängige experimentelle Variablen musikalische Parameter wie Tonskala, melodische Kontur oder Organisation der Töne (z.B. Wiederholung, Sequenzierung, Intervallbeziehungen etc.) zu betrachten, die – so ist zu vermuten – für das Behalten von echten Melodien eine Rolle spielen. Als Beispiel für diese Art von organisierter Tonfolge können die von Ortmann (1933) und später von Taylor und Pembroke (1984) verwendeten 5-Ton-Sequenzen dienen.



Abb. 3.2-1: 5-Ton-Sequenzen als Versuchsstimulus, verwendet von Taylor & Pembroke (1984) bzw. Ortman (1933).

b) Interaktion von Tonhöhenverlauf und Rhythmus: Natürlich ergibt sich erst durch die zeitliche Abfolge von Einzeltönen die Möglichkeit, den Parameter *Rhythmus* als Einflußgröße zu untersuchen, der für das Behalten von echten Melodien unzweifelhaft eine wichtige Rolle spielt (vgl. bspw. Hebert & Peretz, 1997). Für diese Interaktion zwischen Tonhöhenparametern und Rhythmus mögen die verschieden rhythmisierten 5-Ton-Sequenzen von Dowling (1973a) ein anschauliches Beispiel sein.



Abb. 3.2-2: rhythmisierte 5-Ton-Sequenzen als Versuchsstimulus, verwendet von Dowling (1973a), eigene Darstellung.

Im Gegensatz zu Einzeltönen enthalten kurze Tonfolgen also musikalische Parameter, die auch vermutlich für das Behalten von echten Melodien von Bedeutung sind, welches in der Regel das vorrangige Erkenntnisinteresse dieser Studien ist (s.u. Punkt f). Tatsächlich weisen kurze Tonfolgen für das (musik-)psychologische Experimentieren aber auch Vorteile auf, die für ihre Verwendung im Gegensatz zu Melodien aus existierenden Musikstücken sprechen.

c) Unbekanntheit: Im Gegensatz zu Melodien haben kurze Tonfolgen den Vorteil, das sie den Versuchspersonen nicht bekannt sind. D.h. Vorteile in der Erinnerung durch Vorbekanntheit bei einzelnen Vpn. können ausgeschlossen werden, wie dies beispielsweise bei den von Dowling und Fujitani (1971) sogenannten „atonalen“ Sequenzen mit Sicherheit der Fall gewesen ist.



Abb. 3.2-3: „atonale“ 5-Ton-Folgen, verwendet von Dowling & Fujitani (1971), eigene Darstellung.

d) Gute experimentelle Kontrolle: Die verwendeten kurzen Tonfolgen sind in der Regel dadurch charakterisiert, dass sie einige, aber nicht alle musikalischen Parameter mit echten Melodien gemeinsam haben. Dadurch wird erreicht, dass einige Parameter systematisch variiert und als Einflußgrößen untersucht werden können (z.B. Kontur, Tonleiterzugehörigkeit, Intervallik) ohne dass gleichzeitig Veränderungen in anderen Parametern (Rhythmus, Klangfarbe, harmonische Bindung etc.) die Gedächtnisleistung beeinflussen können. Es scheint so möglich, die Bedeutung der wenigen betrachteten Parameter für das Melodiegedächtnis genau evaluieren zu können, ohne an eine ‚musikalische Logik‘ hinsichtlich der Veränderungen in den unterschiedlichen Parametern gebunden zu sein. Für eine wirklich erschöpfend ausgeführte Variation der betrachteten experimentellen Variablen stehen beispielsweise die 3-Ton-Sequenzen von Eiting (1984).



Abb. 3.2-4: variierte 3-Ton-Sequenzen, verwendet von Eiting (1984).

e) Unterbinden von Assoziationen und musikimmanenten Schlussfolgerungen: Implizit wird in den meisten der in diesem Kapitel besprochenen Studien angenommen, dass durch das selektive Ausblenden von verschiedenen musikalischen Parametern und die systematische Variation anderer Parameter ein ‚reines‘ Melodie- oder Tönegedächtnis untersucht werden kann. Die durch die experimentelle Behandlung generierte Künstlichkeit der Tonsequenzen, die sie in der Regel von echten Melodienfragmenten unterscheidet, soll Assoziationen zu bekannten Melodien oder das musikalische Schlussfolgern der Vpn. im Gedächtnistest unterbinden. Die kurzen Tonfolgen stellen somit eine gewisse Analogie zu den Folgen von unverbundenen Zahlen oder Buchstaben dar, wie sie in vielen seriellen Gedächtnisaufgaben für das verbale Gedächtnis verwendet werden. Als bekannte Beispiele können hier genannt werden: die bereits von Ebbinghaus (1885) verwendeten sinnlosen

Silben, die Buchstabentripel des Brown-Peterson-Paradigmas (Brown, 1958) oder die Zahlenreihen der üblichen Gedächtnisspannentests (vgl. Kluwe 1990, S. 142; oder Anderson 1995, S. 169). Stachó & Nemeth (2001) beispielsweise ziehen eine explizite Parallele zwischen den von ihnen verwendeten kurzen Tonfolgen und den genannten Gedächtnistests. Es geht ihnen ausdrücklich darum, das reine Tongedächtnis zu testen, welches ohne Hilfe von gelernten Strukturen aus dem Langzeitgedächtnis auskommt. Gedächtnisleistungen mit diesen Tonfolgen als Stimulusmaterial würden demnach also nicht auf Schlußfolgerungen beruhen, die die Vpn. mittels ihrer musikalischen Erfahrung aus dem Verlauf und der Interaktion aller beteiligten musikalischen Parameter anstellen könnten.

f) Generalisierbarkeit der Ergebnisse: Wie oben erwähnt, ist das hauptsächliche Erkenntnisinteresse, welches hinter den meisten Studien mit kurzen Tonfolgen als Stimulusmaterial steckt, die Beschreibung der Struktur bzw. der Komponenten des Melodiegedächtnisses. Trotz der gerade beschriebenen Unterschiede zu echten Melodien (Selektion nur einiger musikalischer Parameter, Künstlichkeit), werden die Beziehungen der Töne untereinander, die ‚echte‘ Melodien und kurze Tonfolgen gemein haben (z.B. Ableitung aus Tonleiter, Kontur, Organisation etc, s. Punkt a)), zumeist in den Vordergrund gestellt. So erscheint eine Generalisierung der experimentellen Ergebnisse auf den Bereich der echten Melodien möglich und wird zumeist in der Diskussion der Ergebnisse in vielen Studien ohne größere Bedenken angestellt. Zwar wird diese Generalisierung nur sehr selten direkt ausgesprochen und auch meist nicht erwähnt, dass das eigentliche Erkenntnisinteresse dem Gedächtnis für Melodien und nicht den Behaltensleistungen für kurze Tonfolgen gilt. Es werden jedoch häufig die Unterschiede zwischen kurzen Tonfolgen und Melodien sprachlich dadurch verwischt, dass ausschließlich der Terminus „melody“ verwendet wird (vgl. z.B. Dowling 1978; Carterette et al. 1986; DeWitt & Crowder 1986; Mikumo 1992, 1994), obwohl experimentelle Ergebnisse nur zu kurzen Tonfolgen generiert wurden. Eine der wenigen (frühen) Ausnahmen bildet die Studie von Dowling und Fujitani (1971), die in Exp. 1 die Bedeutung der melodischen Kontur für das Behalten von kurzen Tonfolgen und in Exp. 2 für das Wiedererkennen von bekannten Melodien untersucht. In der großen Mehrzahl der empirischen Studien wird jedoch der Geltungsanspruch der Ergebnisse für kurze Tonfolgen relativ subtil auch auf Melodien ausgedehnt, ohne dass dies explizit nachgewiesen würde.

3.2.3 Experimentelle und erkenntnistheoretische Grenzen

Wie gerade beschrieben, lassen sich einige Eigenschaften von kurzen Tonfolgen nennen, die ihre Eignung als Stimulusmaterial für Tests zum Melodiegedächtnis und damit ihre häufige Verwendung in experimentellen Arbeiten zum Melodiegedächtnis erklären. Zum Teil sind es jedoch gerade diese Eigenschaften, die auch für die Grenzen des Geltungsanspruches der Ergebnisse in diesen Arbeiten verantwortlich sind und auf die in jüngerer Zeit einige Autoren aufmerksam gemacht haben.

Ausblenden von subjektabhängigen Variablen: Wie Gillian Cohen (1998) bemerkt, ist das Gedächtnis für Musik und für Melodien, wie es unter alltäglichen Bedingungen funktioniert, nicht nur abhängig von Parametern der musikalischen Struktur selbst, sondern auch von den musikalischen Erfahrungen des Subjekts. Dazu zählt er in Anlehnung an eine Studie von Russell (1987) die Musikpräferenzen und die Vertrautheit mit einem musikalischen Stil (Cohen 1998, S. 218). In den Studien mit kurzen Tonfolgen als Stimulusmaterial werden diese Variablen jedoch bewusst ausgeblendet; d.h. sie werden erstens nicht erhoben und nicht in die Datenanalyse einbezogen, und zweitens wird auch bei der Konstruktion der kurzen Tonfolgen darauf geachtet, dass die kurzen Tonfolgen neutral und frei von Ähnlichkeiten zu echten melodischen Floskeln bleiben. Über den Einfluß der personenbezogenen Variablen wie Präferenzen und der Vertrautheit mit einem Stil können Experimente mit kurzen Tonfolgen also keine Auskunft geben. Nicht zuletzt aus diesem Grund kommt Cohen zu der ernüchternden Feststellung:

Memory for music is an area where traditional laboratory testing has dominated the research and the results may not be very relevant to the ability of the average untrained person to remember music in natural everyday situations. (Cohen 1998, S. 220)

Vernachlässigung von formalen melodischen Gliederungen: Aus kurzen Tonfolgen sind in der Regel keine melodischen Bauprinzipien zu entnehmen, die für die Konstruktion und die Analyse von größeren melodischen Einheiten bestimmend sind. Dazu gehören die Bildung von melodischen Abschnitten, Abschnittswiederholungen, Kadenzen oder größere melodische Formeln. Es sind gerade diese formbildenden Strukturen, die die Vpn. in einem Versuch von Sloboda und Parker (1985), der einen Ausschnitt von 30 Noten aus einer echten Melodie als Versuchsmaterial verwendet, besonders gut erinnern und wiedergeben konnten. Wie Sloboda und Parker bemerken, sind diese strukturellen Variablen in ihrer Bedeutung für das musikalische Gedächtnis nur zu beobachten, wenn das melodische Testmaterial ‚echten‘ Melodien hinreichend ähnlich ist und eine gewisse Länge hat (Slo-

boda & Parker 1985, S. 146). Sloboda und Parker vermuten nach den Ergebnissen ihrer Untersuchung, dass das Memorieren von Melodien zuerst darin besteht, die mentale Struktur einer Melodie aufgrund von formalen Informationen aufzubauen, mit deren Hilfe dann musikalische Details der Melodie in einer Art rekonstruktivem Prozess erschlossen werden. Diese These wird auch von Cohen am Ende seiner Literaturdurchsicht gestützt, wenn er von „representational and reconstructive processes“ (Cohen 1998, S. 220) spricht, die Personen bei der Erinnerung an Musik einsetzen. Aufgrund der Wirkung dieser rekonstruktiven Prozesse weist Edwin Hantz darauf hin, dass es möglich ist, ganz unterschiedliche und u.U. widersprüchliche Behaltensleistungen für Teile einer Melodie und für ein Melodieganzes zu erhalten. Diese Möglichkeit wird jedoch bei der Generalisierung von Ergebnissen für kurze Tonfolgen in der Regel nicht in Betracht gezogen (Hantz, 1984).

Vernachlässigung von melodischen Schemata: Für die Wahrnehmung und Erinnerung von melodischen Gestalten in echter Musik scheinen auch melodische Schemata von einiger Bedeutung zu sein, wie Bob Snyder in Anschluß an Leonard Meyer ausführt (Snyder 2000, S. 152-155). Die Schemata sind Muster von melodischen Folgen, die entweder innerhalb von spezifischen Musikstilen gelernt werden oder als bildhafte Analogien quasi intermodal verstanden werden. Die in den meisten hier in Frage stehenden Experimenten verwendeten kurzen Tonfolgen sind jedoch zu kurz und im oben beschriebenen Sinne zu künstlich, um als melodisches Schema verstanden und erinnert zu werden. Die rekonstruktiven Prozesse, die auf melodischen Schemata beruhen, werden deshalb auch in den meisten Versuchen mit kurzen Tonfolgen nicht beachtet.

Zusammenfassend lässt sich von der vorangegangenen Erörterung bezüglich Eignung und Geltungsanspruch von kurzen Tonfolgen als Versuchsmaterial für Untersuchungen zum Melodiegedächtnis folgendes festhalten:

Kurze Tonfolgen scheinen gerade für mehrfaktorielle varianzanalytische Pläne als Versuchsmaterial gut geeignet, da hier eine möglichst vollständige und systematische Variation entlang der untersuchten musikalischen Dimensionen von großer Bedeutung ist (vgl. Borz 1999, S. 279-280). Dies ist mit kurzen Tonfolgen, die noch keinen Regeln der Melodiebildung unterworfen sind, einfacher zu erreichen als mit echten Melodien, die ihren musikalischen Charakter durch eine systematische Variation eines Parameters verlieren würden. Die resultierende Künstlichkeit und der qualitative Unterschied zu echten Melodien werden dabei sogar eher als Vorteil erachtet, denn von der Assoziierbarkeit zu musi-

kalischem Vorwissen und zu Inhalten im LZG sollen die experimentellen Ergebnisse unbeeinflusst bleiben.

Gleichzeitig scheint aber gerade diese Künstlichkeit und der ‚unmusikalische‘ Charakter der kurzen Tonfolgen die Generalisierbarkeit der experimentellen Ergebnisse auf den Bereich der echten Melodien zu verhindern, wie einige Autoren in der jüngeren Vergangenheit dargelegt haben. Monahan et al. (1987) machen auf dieses Dilemma zwischen der Kontrolle über die experimentellen Variablen und die Generalisierbarkeit der Ergebnisse auf ‚echte‘ Musik aufmerksam (S. 599) und beschreiten selbst einen Mittelweg (s. die Diskussion von Monahans Akzentmodell im folgenden Unterkapitel). Für die große Mehrheit von Studien gilt jedoch, dass wichtige Parameter für die Wahrnehmung von melodischer Gestalt und die musikalische Erinnerungsfähigkeit mit kurzen Tonfolgen als Versuchsmaterial nicht abgeprüft werden können. Zu diesen vernachlässigten Parametern gehören musikalische Präferenzen, die Vertrautheit mit einem musikalischen Stil, die formale melodische Anlage, die Interaktion zwischen den verschiedenen musikalischen Parametern sowie die Bedeutung von Melodieschemata, die den Versuchspersonen bereits bekannt sind. Wie Kauffman und Carlsen feststellen, fehlt den in diesen Studien verwendeten kurzen Tonfolgen die ökologische Validität, die sie mit echter Musik vergleichbar machen würde:

The principal ingredient lacking in these as well as in similar studies [Studien von Ortmann, Taylor & Pembroke und Dowling] is the ecological validity for music as musicians know it. (Kauffmann & Carlsen 1989, S. 3)

Auffällig ist, dass trotz dieses offensichtlichen Mankos, welches die vielen Studien aufweisen, die kurze Tonfolgen als Stimulusmaterial verwenden, und trotz der weiten Verbreitung der Ergebnisse aus diesen Studien in der Literatur bisher kein ausführlicher methodischer, epistemologischer oder empirischer Nachweis versucht wurde, der die Generalisierung der bisherigen Ergebnisse für den Bereich der Melodien oder ‚echter‘ Musik rechtfertigen würde. Wahrscheinlich ist es nicht zuletzt dem Fehlen dieses Nachweises und damit den Zweifeln am Geltungsbereich der experimentellen Ergebnisse, die mit kurzen Tonfolgen entstanden sind, zuzuschreiben, dass in den letzten 15 Jahren eine Reihe von Studien entstanden, die statt kurzer Tonfolgen echte Melodien für Experimente zum Melodiegedächtnis verwenden. Als Beispiel für diese Hinwendung zu Stimulusmaterialien, die eine größere ökologische Validität aufweisen, mögen die Studien von Walter Jay Dowling selbst gelten, der die abstrakten, künstlichen Stimuli seiner Versuchsreihen der 70er und 80er Jahre zu Gunsten ökologischer Validität und Einbeziehung weiterer relevanter Variablen verließ. Die von Dowling und Mitarbeitern später fokussierten Variablen waren beispielsweise harmonischer Kontext und echte Instrumentalklangfarbe (Dowling, 1986), to-

nale Beziehungen und Rhythmen in Ausschnitten aus vorbestehenden Melodien (Dowling et al., 1995, Exp. 5-7) und Phrasen aus existenten Musikstücken im richtigen musikalischen Kontext (Dowling et al., 2002). Dowling selber spricht explizit von größerer ökologischer Validität (Dowling et al., 2002, S. 254), die durch die Verwendung von Ausschnitten aus echter Musik erreicht wird.

3.2.4 Untersuchte Variablen und experimentelle Paradigmen

3.2.4.1 Experimentelles Paradigma

Das Erkenntnisinteresse der meisten hier angeführten Studien besteht – wie oben bereits kurz dargestellt – darin, die für das Behalten von Melodien wichtigen musikalischen Dimensionen zu identifizieren. Um in diesem Sinne zu Ergebnissen zu gelangen, wird häufig ein Experimentaldesign verwendet, welches sich aus folgender Überlegung ableitet: Wird eine kurze Tonfolge hinsichtlich einer musikalischen Dimension – d.h. Variable (z.B. melodische Kontur oder Rhythmus) – über eine gewisse Zeit ohne Fehler im Gedächtnis behalten, dann scheint diese Dimension für die musikalische Wahrnehmung bzw. für das melodische Gedächtnis von Bedeutung zu sein.

Diese Vorhersage lässt sich prinzipiell auf verschiedene Weise experimentell überprüfen. Jedoch hat sich in der Forschung mit kurzen Tonfolgen vor allem das Paradigma des Wiedererkennens (*recognition*) und des Vergleichs von zwei Stimuli durchgesetzt (zum Wiedererkennen-Paradigma s.a. Kluwe, 1990, S. 122ff). Zwar existieren einige Studien, die an Stelle des Wiedererkennens- ein Wiedergabeparadigma verwenden (z.B. Roberts et al., 1983; Taylor & Pembroke, 1984; Roberts, 1986; Cutietta & Booth, 1996). Doch insgesamt überwiegt in der Literatur deutlich das Wiedererkennen als Testmethode; wahrscheinlich aufgrund der bei der Wiedergabe schwierig zu erhebenden abhängigen Variablen (s. Diskussion dieses Punktes bei Sloboda & Parker (1985) und die Diskussion dieser Studie im folgenden Unterkapitel).

Der Prototyp eines Experimentes zum Wiedererkennensgedächtnis für kurze Tonfolgen lässt sich wie folgt beschreiben:

Der Versuchsperson wird eine kurze Tonfolge vorgespielt und nach einer kurzen Pause (Retentionsintervall) wird ihr eine zweite kurze Tonfolge präsentiert, die entweder identisch zur ersten Tonfolge sein kann oder sich von dieser hinsichtlich einer oder mehrerer Variablen unterscheidet. Die Versuchsperson muss nun angeben, ob die zweite Tonfolge

zur ersten identisch ist oder nicht. Dabei wird angenommen, dass die Versuchsperson die zweite Tonfolge mit der Erinnerung an die erste Tonfolge im Gedächtnis vergleicht.

Bedeutung für das musikalische Gedächtnis hat eine Variable dann, wenn Unterschiede bzw. Identitäten zwischen den beiden zu vergleichenden Melodien in Hinblick auf diese Variable signifikant häufig von den Versuchspersonen richtig erkannt werden (vgl. auch die Erläuterung dieses Gedankens bei Cohen, 1998, S. 218). Die Schlussfolgerung aus diesem Ergebnis ist, dass die untersuchte Variable, die eine bestimmte Eigenschaft der Tonfolge repräsentiert, richtig im Gedächtnis behalten wurde und so ein korrektes Urteil hinsichtlich der Identität der beiden Tonfolgen möglich wurde. Da die allermeisten der im folgenden besprochenen Studien diesem experimentellen Wiedererkennensparadigma folgen, soll hier noch die knappe und sehr verständliche Zusammenfassung dieses Ansatzes von Idson und Massaro (1978) angeführt werden:

A melody recognition task was employed, in which the melodies were subjected to a variety of structural transformations designed to preserve or violate selective sources of information. The level of performance on these transformed melodies, with respect to the level of performance on the original untransformed melodies, provides an index of the importance of the manipulated information. To the extent that distortion of a source of information produces a performance decrement, that information can be thought to play a role in melodic perception. Similarly, if a given source of information can be eliminated without disrupting performance, then that information is not necessary for accurate melodic perception. (Idson & Massaro, 1978, S. 554)

Die Möglichkeit, dass eine Tonfolge korrekt im Gedächtnis behalten wird und die Versuchsperson die Identität der beiden Tonfolgen dennoch nicht korrekt beurteilt, wurde generell in der experimentellen Literatur bisher nicht in Betracht gezogen. Wie jedoch beispielsweise die Studie zur Reproduktion von Intervallen mit Hilfe eines Oszillators von Attneave und Olson (1971, Exp. 1) zeigt, haben Personen ohne musikalische Vorbildung erhebliche Schwierigkeiten, größere musikalische Intervalle transponiert, auf einer anderen absoluten Tonhöhe korrekt nachzubilden, auch wenn das Referenzintervall ständig wiederholt werden kann, d.h. das Gedächtnis nur wenig beansprucht wird. Dies gilt zumindest für künstliche Tonfolgen (dort: einzelne Intervalle), die mit nicht-musikalischen Klangfarben erzeugt werden (dort: Sinustöne).

Das beschriebene Experimentaldesign legt eine varianzanalytische statistische Auswertung nahe, denn verschiedene Zustände oder Ausprägungen einer musikalischen Dimension (d.h. Variablen) werden auf ihre Wirkung auf das Behalten der Tonfolgen untersucht. Die Ausprägungen der musikalischen Variablen stellen in der Regel nominale Kategorien dar, während die Behaltensleistung der Versuchspersonen in mindestens einer metrischen Vari-

able erfasst wird. Es erscheint hier deshalb sinnvoll, eine kurze Beschreibung der in der Vergangenheit untersuchten abhängigen und unabhängigen Variablen folgen zu lassen, die den Untersuchungsrahmen dieser Forschungen festsetzen.

In mehrfaktoriellen Designs stellen die Ausprägungen einer Variable Bedingungen dar, unter denen die Bedeutsamkeit einer (oder mehrerer) anderer Variablen für die Behaltensleistung geprüft werden. Von Bedeutung sind dabei auch Interaktionseffekte zwischen diesen Variablen, die anzeigen, dass eine Variable unter gewissen Bedingungen für das Behalten von kurzen Tonfolgen von Bedeutung ist, unter anderen Bedingungen aber nicht.

3.2.4.2 Abhängige Variablen

Die abhängige Variable stellen bei diesen Wiedererkennenstests die Urteile der Versuchspersonen dar. Zur Angabe der Identität der beiden Testsequenzen stehen der Versuchsperson im einfachsten Fall die Kategorien „Ja“ und „Nein“ (identisch / nicht-identisch) zur Verfügung. In einigen Varianten dieses Paradigmas kann die Versuchsperson zusätzlich noch die Sicherheit des eigenen Urteils auf einer ordinalen Skala (i.d.R. 3-5 Stufen) angeben. Aus diesen dichotomen Ja/Nein-Daten kann durch einen Vergleich mit der tatsächlichen Identität bzw. Nicht-Identität der beiden Stimuli eine dichotome Variable bestimmt werden, die die richtigen Antworten der Versuchsperson wiedergibt. Die Antworten aller Versuchspersonen zu einem *Testitem* werden anschließend zusammengezählt. Daraus ergibt sich dann eine metrische Variable „richtige Antwort“, die entweder die Anzahl aller richtigen (bzw. falschen) Antworten aller Versuchspersonen zum jeweiligen *item* oder alternativ den Anteil der richtigen (bzw. falschen) Antworten an allen Antworten zu einem *Testitem* wiedergibt. Diese metrische Variable wird in der Regel als abhängige Variable für die anschließenden varianzanalytischen Auswertungen verwendet. Statt der „richtigen Antworten“ wird in einigen Studien das Maß d' aus der Signal-Detektionstheorie verwendet, welches die individuelle Tendenz einer Versuchsperson, eher „Ja“ oder eher „Nein“ zu antworten, ‚herausrechnet‘ und so neutralisiert (s. Exkurs).

Eine etwas differenziertere Darstellung liefert die gleichzeitige Erhebung von Urteilssicherheit und richtiger Antwort in einer einzigen Variablen, die vier, fünf oder sechs ordinale Stufen verwendet (z.B. „sure yes“, „yes“, „no“, „sure no“). Hieraus bilden Dowling und einige andere Forscher die sogenannte „area under the memory operating characteristic“ (MOC) – eine an der Signal-Detektionstheorie angelehnte Repräsentation von Gedächtnisleistungen – und das zugehörige Maß AUC . Bei diesem Maß werden die

Ratingwerte für korrekte „Treffer“ und „falsche Alarme“ gegeneinander verrechnet. Das *AUC*-Maß wurde v.a. von W. Jay Dowling (1972, 1973a, 1978, 1991), Dowling und Fujitani (1971), Long (1977), aber auch von DeWitt & Crowder (1986) und Monahan et al. (1987) verwendet.

3.2.4.2.1 Exkurs: Die Maße d' und *AUC*

Obwohl Details zur Methodik und zur Verwendung von d' und der MOC bzw. des *AUC*-Maßes in Gedächtnisexperimenten (bzw. zu den analog konstruierten sogenannten ROC-Kurven) sich in vielen Studien finden (z.B. Norman & Wickelgren, 1965; Green & Moses, 1966; Pollack & Hsieh, 1969; Banks, 1970; Swets, 1973), soll in diesem Exkurs in aller Kürze die Konstruktion der beiden Maße erläutert werden. Zum einen hat dieser Exkurs einen didaktischen Zweck, da die beiden Maße zumindest in der deutschen Musikpsychologie sehr wenig verwendet werden und wahrscheinlich weitgehend unbekannt sind. Zum anderen soll auch auf die hinter den beiden Maßen stehende experimentelle Logik und die Konsequenzen für das Versuchsdesign und damit auch die Art der Erkenntnisse eingegangen werden, die unter Verwendung dieses Ansatzes für das Melodiegedächtnis gewonnen werden können.

Die *Signal Detection Theory* (SDT) wurde in den vierziger Jahren des vergangenen Jahrhunderts zuerst im militärischen Bereich entwickelt. An ihrem Anfang stand die Frage, wie die Performanz von Systemen oder menschlichen Beobachtern zu messen sei, die aus einem rauschbehafteten Übertragungskanal Signale entdecken sollen. In den sechziger Jahren fand eine Adaption für die psychologische und psychophysische Forschung statt, innerhalb derer die Maße der SDT im auditorischen Bereich vor allem zur Bestimmung von Hörschwellen und Unterscheidungsleistungen eingesetzt wurden (vgl. MacMillan & Creelman, 1991). Eine einflussreiche Veröffentlichung hierzu war das Buch von David Green und John Swets *Signal Detection Theory and Psychophysics* (1966). Das Ziel bei der Verwendung des SDT-Ansatzes im psychologischen Experiment war, „die motivationalen Komponenten von den sensorischen zu trennen“ (Irtel, 1996, S. 482); d.h. die individuelle Tendenz einer Versuchsperson, einer bestimmten Antwortkategorie zuzuneigen, ‚herauszurechnen‘ und so die ‚reine‘ psychophysische Leistungsfähigkeit (z.B. die Hörschwelle) zu bestimmen.

Einer Versuchsperson werden innerhalb eines Experimentes also Reize in derselben experimentellen Bedingung mehrfach dargeboten und die Antworten aufgezeichnet. Bei einem Experiment, bei dem es nur zwei Antwortkategorien gibt (z.B. ‚Ja, Ton gehört‘ und ‚Nein, kein Ton gehört‘) sind damit vier Ergebnisse denkbar, die in der folgenden kleinen Tabelle mit den gebräuchlichen englischen Begriffen bezeichnet sind. Die Zahlen in Klammern sind Werte eines fiktiven Beispiels, das dem Leser den Nachvollzug durch die Möglichkeit des Nachrechnens erleichtern soll.

	Antwort der Vpn.	
Stimulus	Ja (24)	Nein (36)
Präsent (30)	Hits (20)	Misses (10)
Nicht präsent (30)	False alarms (4)	Correct rejections (26)

Tabelle 3.2–1: Antwortkategorien in *Signal-Detection*-Experimenten; fiktive Beispielwerte in Klammern.

Es wird nun der Anteil der *hits* an allen Versuchsdurchgängen bestimmt, in denen der Stimulus präsent war ($H = 20/30 = 0,67$) sowie der Anteil der *false alarms* an allen Durchgängen, in denen der Stimulus nicht präsent war ($F = 4/30 = 0,13$). Da die *hits* zwar für eine gute Erkennensleistung sprechen, die *false alarms* aber dagegen, ist es sinnvoll, ein Maß zu finden, das beide Kategorien berücksichtigt. Dies ist z.B. das Maß d' , bei dem der Anteil der *false alarms* vom Anteil der *hits* abgezogen wird. Um normierte Werte zu erhalten, werden nicht die rohen Verhältnisse von einander abgezogen, sondern die entsprechenden Werte der inversen Dichtefunktion der Normalverteilung (hier mit z benannt). H und F werden als Flächen unter einer Standardnormalverteilungskurve gesehen, deren zugehörige Werte auf der Abzisse (z -Werte) werden von einander abgezogen. So errechnet sich d' wie folgt:

$$d' = z(H) - z(F)$$

Mit den Werten für $z(0,67) = 0,44$ und für $z(0,13) = -1,12$ des obigen Beispiels:

$$d' = 0,44 - (-1,12) = 1,56$$

Ein Wert von $d' = 0$ zeigt an, dass die Vpn. keine Erkennensleistung in dieser Versuchsbedingung besitzt; je größer die Werte von d' , desto besser auch die Unterscheidungsleistung der Vpn. hinsichtlich der Versuchsbedingung.

Sind die Antworten der Vpn. nicht „Ja“ und „Nein“, sondern die Werte einer *Ratingskala*, so lässt sich für jeden Skalenwert zu jeder Versuchsperson in jeder Bedingung ein d' -Wert berechnen. Als fiktives Beispiel sollen die Werte der folgenden Tabelle dienen, die zu einer gewissen Experimentalbedingung einer Versuchsperson gehören, in der ein Stimulus

präsent oder nicht präsent sein konnte. In Klammern sind die Verhältnis jeder Zelle zum Gesamtwert der Zeile angegeben:

	Sicher ja	Ja	Weiss nicht	Nein	Sicher nein	Gesamt
Stimulus präsent	14 (0,28)	12 (0,24)	11 (0,22)	8 (0,16)	5 (0,1)	50 (1)
Stimulus nicht präsent	6 (0,12)	9 (0,18)	13 (0,26)	11 (0,22)	11 (0,22)	50 (1)

Tabelle 3.2-2: Fiktive Beispielwerte für die *ratings* einer Vpn. Auf einer fünfstufigen Skala in einem Wiedererkennensexperiment. Zugehörige Verhältnisse zu allen *ratings* einer Zeile in Klammern.

Als nächstes werden die Verhältniswerte der Zellen in jeder Zeile von links nach rechts kumuliert und die kumulierten Werte z-transformiert (Werte in Klammern in der folgenden Tabelle). Für jeden Wert der *Ratingskala* steht nun also ein *H*- und ein *F*-Wert bzw. ein $z(H)$ - und ein $z(F)$ -Wert bereit. Schließlich werden d' -Werte für jede Spalte, d.h. für jeden Werte der *Ratingskala*, berechnet, die in der untersten Zeile der folgenden Tabelle aufgeführt sind.

	Sicher ja	Ja	Weiss nicht	Nein	Sicher nein
Stimulus präsent	0,28 (-0,58)	0,52 (0,05)	0,74 (0,65)	0,9 (1,29)	1
Stimulus nicht präsent	0,12 (-1,17)	0,3 (-0,52)	0,56 (0,15)	0,78 (0,77)	1
d'	0,59	0,57	0,5	0,52	

Tabelle 3.2-3: Von links nach rechts kumulierte Verhältnisse aus Tabelle 3.2-2 und zugehörige z-Werte in Klammern.

Die zusammengehörenden Werte von *H* und *F* lassen sich als Punkte in einem zweidimensionalen Raum darstellen, der durch die Achsen *hit*-Rate und *false alarms*-Rate aufgespannt wird. Verbindet man diese Punkte zu einer Kurve, erhält man die sogenannte *Receiver Operating Characteristic* (ROC – in Gedächtnisexperimenten oft *Memory Operating Characteristic* (MOC) genannt), die eine Darstellung der Leistungsfähigkeit der Vpn. in einer Versuchsbedingung darstellt. Die Kurve, die sich aus den Werten des obigen Beispiels ergibt, ist in Abbildung 3.2-5 dargestellt.

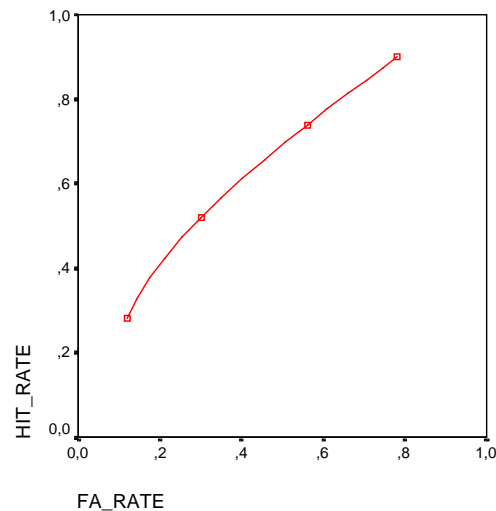


Abb. 3.2-5: MOC-Kurve des fiktiven Beispiels aus Tabelle 3.2-3

Die gebräuchlichste Methode, diese Kurve auf einen einzigen Wert zu reduzieren, ist die Fläche unter der Kurve zu berechnen, die zusätzlich vom Ursprung des Koordinatensystem bis zum Wert (1,1) extrapoliert wird. Dies ist das sogenannte *AUC*-Maß (Area under Curve). Sowohl zur Verbindung der empirischen *hit*- und *false alarms*-Anteile zu einer Kurve sowie zu deren Extrapolation als auch zur Berechnung der Fläche werden in den verschiedenen Software-Paketen unterschiedliche Methoden verwendet, die aber generell zu sehr ähnlichen Ergebnissen führen. Die extrapolierte Kurve mit der Diagonale als Vergleichslinie sei schließlich noch als SPSS-Grafik dargestellt. Die *y*-Achse wird dort als *Sensitivität* bezeichnet, also als Fähigkeit, den experimentellen Stimulus zu erkennen (richtigerweise mit ‚ja‘ zu antworten). Die *x*-Achse misst die *Spezifität* (genauer *1 - Spezifität*), d.h. die Fähigkeit den Stimulus von Rauschen oder anderen Stimuli zu unterscheiden (richtigerweise mit ‚nein‘ zu antworten).

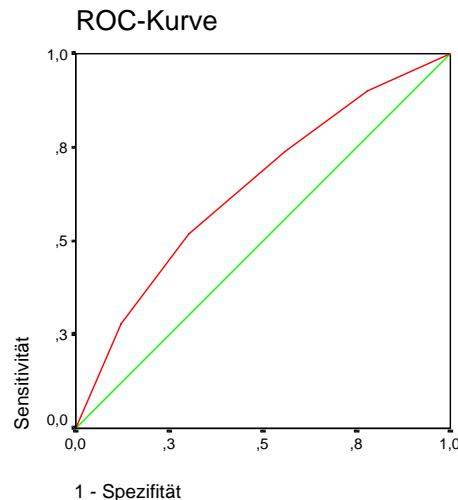


Abb. 3.2-6: In SPSS berechnete MOC-Kurve (in SPSS ROC-Kurve genannt) des fiktiven Beispiels aus Tabelle 3.2-3 mit Diagonale als Vergleichslinie.

SPSS errechnet als Fläche unter der Kurve den *AUC*-Wert 0,642. Die Fläche unter der Diagonale ist dagegen 0,5. Dieser *AUC*-Wert steht für keinerlei Unterscheidungsleistung. Wölbt die MOC-Kurve sich in Richtung des Werte (0,1), d.h. zur ‚linken oberen Ecke‘, so steht dies für eine bessere Unterscheidungsleistung. Ein *AUC*-Wert von 1 steht für eine perfekte Unterscheidungsleistung einer Vpn. Nach der Klassifikation von Hosmer und Lemeshow (2000) repräsentieren *AUC*-Werte von $<0,7$ keine akzeptable Unterscheidungsleistung. Die fiktive Versuchsperson dieses Beispiels schnitt also in dem fiktiven Experiment schlecht ab.

Essentiell für die Berechnung sowohl von d' als auch des *AUC*-Maßes ist, dass genug Durchläufe einer Vpn. unter jeder Versuchsbedingung vorgenommen werden, d.h. dass genug Daten zur Verfügung stehen. Da die Anzahl der Antworten in den vier Kategorien der obigen Tabelle gegeneinander verrechnet werden, müssen mindestens zwei Antworten pro Bedingung (eine für Stimulus präsent, eine für Stimulus nicht-präsent) vorliegen, um überhaupt relative Anteile bilden zu können. Entwickelt wurden die Maße aber für die Betrachtung einer Vielzahl von Antworten unter derselben Bedingung. Es ist leicht nachzuvollziehen, dass in einem Gedächtnisexperiment, welches diese Maße verwendet, nicht mehr als 18-24 unterschiedliche Faktorstufenkombinationen mit einer Versuchsperson getestet werden können. Dies entspricht z.B. einem 4x3x2 Faktorstufendesign. Mehr als 48 unterschiedliche *items* (Tonfolgenpaare) kann eine Vpn. (pro Experimentalsession) schon aus Konzentrationsgründen nicht mit der gleichen Qualität beurteilen. Es können also nur

wenige Faktoren, die für das Melodiegedächtnis eine Rolle spielen könnten, in einem solchen SDT-Design gleichzeitig betrachtet werden.

Alle anderen Eigenschaften des Testmaterials, die sich z.B. zufällig ergeben, müssen als ‚Rauschen‘ (‚noise‘) betrachtet werden und können deshalb nicht berücksichtigt werden. Die Wahl der SDT-Maße als abhängige Variablen bedingt also eine simplifizierte Betrachtung des melodischen Testmaterials. Eine Reduktion der Qualitäten einer kurzen Tonfolge oder einer Melodie auf zwei bis drei wichtige Merkmale mit je zwei bis vier relevanten Ausprägungen ist unumgänglich.

3.2.4.2.2 Kritik der für die abhängigen Variablen verwendeten Maße

Einer der Gründe für die weite Verbreitung des Wiedererkennensparadigmas in Experimenten mit kurzen Tonfolgen und bei Versuchen zum musikalischen Gedächtnis überhaupt ist die einfache Handhabbarkeit der von den Versuchspersonen erhobenen Daten, d.h. der abhängigen Variablen als Anteil der korrekten Antworten bzw. als d' oder AUC . Während Gedächtnistests, in denen sprachliches Versuchsmaterial wiedergegeben werden muss, relativ einfach auf Fehler und richtige Antworten auszuwerten sind, ist dies bei musikalischen Stimuli sehr viel schwieriger. Werden musikalische Elemente von den Versuchspersonen wiedergegeben, dann sind die Ergebnisse fast zwangsläufig ‚schmutzige‘ Daten, d.h. gesungene oder auf einem Instrument gespielte Tonfolgen, die in aller Regel nicht exakt ausgeführt werden, sondern in mehrerer Hinsicht musikalisch ungenau sind. Eine objektive Zuordnung dieser Ungenauigkeiten zu den binären Kategorien ‚richtig‘ oder ‚falsch‘ gestaltet sich bei der Datenauswertung schwierig.

Zusätzlich ist bei der Auswertung dieser ‚schmutzigen‘ Daten im Nachhinein schwer zu entscheiden, ob die Ungenauigkeiten primär auf Fehler im musikalischen Gedächtnis der Versuchspersonen zurückzuführen sind oder lediglich auf Fehler bei der musikalischen Produktion.

In Wiedererkennenstests wird von den Versuchspersonen lediglich verlangt, Urteile in vorgegebenen Kategorien abzugeben. Diese Urteile sind entweder als richtig oder als falsch auszuwerten (s.o.) und der Experimentator kann somit ohne großen Aufwand präzise Daten für die abhängige Variable erhalten, die die statistischen Voraussetzungen für die Überprüfung der formulierten Hypothesen erfüllen. Das Problem der Bewertung ungenauer musikalischer Wiedergaben besteht im Wiedererkennensparadigma nicht.

Wie Sloboda und Parker (1985) jedoch ausführen, lassen sich durch das Wiedererkennensparadigma oft die differenzierteren Mechanismen musikalischer Verarbeitung und musikalischen Gedächtnisses durch die Reduktion auf zwei oder wenige mögliche Antwortkategorien nicht überprüfen.

In many cases, the binary theoretical choices that this forces one to make are naïve and ill-conceived. (Sloboda & Parker, 1985, S. 146)

Darüber hinaus bemerken Sloboda und Parker, dass die Aufgaben, die den Versuchspersonen in Wiedererkennensexperimenten mit kurzen Tonfolgen gestellt werden, selten einem normalen musikalischen Verhalten entsprechen. Aufgaben, wie das Erkennen der melodischen Kontur in chromatischen Tonfolgen (Dowling & Fujitani, 1971, Exp. 1) oder die Identifizierung von kurzen chromatischen Sequenzen in Umkehrung, Krebs oder Krebsumkehrung (Dowling 1972) werden von Sloboda und Parker zwar nicht angeführt, können aber sicher als Beispiel in diesem Sinne gelten. Nicht nur die Erhebung der abhängigen Variable, sondern auch das experimentelle Design an sich, wird nach der Ansicht von Sloboda und Parker in vielen Studien den dahinterliegenden Fragen nach der Funktionsweise des musikalischen Gedächtnisses, wie es in alltäglichen Situationen arbeitet, nicht gerecht.

3.2.4.3 Unabhängige Variablen

Im Vordergrund des Interesses stehen bei den hier besprochenen Experimenten mit kurzen Tonfolgen die verschiedenen musikalischen Eigenschaften, die die Versuchspersonen quasi von der Tonfolge abstrahieren und im Gedächtnis behalten bzw. durch welche die Tonfolge an sich erinnerbar wird. So ist es wenig verwunderlich, dass fast alle wichtigen Dimensionen, die kurze Tonfolgen aufweisen können, bereits als unabhängige Variablen experimentell untersucht wurden.

Ob innerhalb einer musikalischen Dimension eine bestimmte Eigenschaft einer Tonfolge richtig im Gedächtnis behalten wird, wird dabei dadurch geprüft, dass untersucht wird, ob Unterschiede hinsichtlich dieser Eigenschaft zwischen den beiden Tonfolgen richtig erkannt werden (s.o.). Das bedeutet, dass die Art der Manipulation einer musikalischen Dimension (von der ersten zur zweiten Tonfolge) den Faktor im varianzanalytischen Sinne bildet. Die einzelnen Faktorstufen stellen dabei die verschiedenen Manipulationsarten dar, die sich jeweils auf eine Eigenschaft innerhalb dieser Dimension beziehen.

In diesem Zusammenhang ist der Tonhöhenverlauf eine der am häufigsten untersuchten Dimensionen. Zu dieser Dimension sind alle Eigenschaften zu rechnen, die sich als Beziehungen zwischen den aufeinanderfolgenden Tönen herstellen lassen. Dazu gehören die ge-

nerelle Bewegungsrichtung der Tonfolge, die melodische Kontur (das Muster von auf- und absteigenden sowie gleichbleibenden Tonhöhen), die absolute Tonhöhe, die relative Größe der aufeinanderfolgenden Intervalle und die exakte Intervallgröße. Diese verschiedenen Eigenschaften unter der Dimension Tonhöhenverlauf bzw. die Manipulationen dieser Eigenschaften bei einer gegebenen Tonfolge sind die Faktorstufen, die in ihrer Bedeutung für das Behalten einer kurzen Tonfolge untersucht werden.

Als einflussreiches Beispiel lässt sich die Studie von Dowling und Fujitani (1971) anführen, die die Bedeutung der melodischen Kontur für das Wiedererkennen von kurzen Tonfolgen herausstellt. Die Faktorstufen des ersten Faktors im Experiment 1 dieser Studie waren: Manipulation der Intervalle (bei erhaltener Kontur), gleichzeitige Manipulation von Intervallen und Kontur, Manipulation von Kontur oder von Kontur & Intervallen (ohne identische Vergleichstonfolgen). Der zweite Faktor war die Transposition der Vergleichstonfolge gegenüber der ersten Tonfolge mit den beiden Faktorstufen ‚transponiert‘ und ‚nicht transponiert‘. Dowling und Fujitani gehen in ihrer Studie von der Bedeutung der musikalischen Kontur für das Wiedererkennen von melodischen Motiven in ‚echter‘ Musik aus. Als Beispiel führen sie in der Einleitung des Textes die Wiederholungen des Themas im Verlauf einer Fuge (Comes) an, welche so gut wie immer die melodische Kontur erhalten, die einzelnen Intervalle jedoch der zugrundeliegenden Skala bzw. den Tonstufen anpassen. Offenbar reiche für das Wiedererkennen eines Fugenthemas aber die Konturinformation aus, was mit dem ersten Experiment ihrer Studien anhand von kurzen Tonfolgen belegt werden soll. Die Kontur einer kurzen Tonfolge wird dabei allein durch die Richtung ihrer Intervalle definiert, wobei lediglich die drei Werte „auf“, „ab“ und „gleichbleibend“ vorkommen können.

Die melodische Kontur wurde in der Folge auch in vielen weiteren Studien mit anderen Eigenschaften der Dimension Tonhöhenverlauf in ihrer Bedeutung für das Behalten von kurzen Tonfolgen verglichen (z.B. Cuddy et al., 1981; Cuddy & Lyons, 1981; Edworthy, 1983; Taylor & Pembroke, 1984; Carterette et al., 1986; Lindsay, 1996).

Neben den Eigenschaften des Tonhöhenverlaufs wurden in einigen, zahlenmäßig geringeren Studien weitere musikalische Parameter wie Rhythmus (z.B. Dowling, 1973a; Monahan et al., 1987), harmonisches Gerüst oder Anlehnung an harmonische Strukturen bzw. Tonalität (z.B. Cohen et al., 1989, Dowling, 1991; Dowling et al., 1995) oder formaler Aufbau der Tonfolge (z.B. Deutsch, 1980; Monahan & Carterette, 1985; Boltz & Jones, 1986;) untersucht.

Neben diesen Variablen, die die musikalischen Eigenschaften der experimentellen Stimuli betreffen, war – fast seit dem Beginn der Forschung mit kurzen Tonfolgen – auch das Niveau der musikalischen Expertise der Versuchspersonen oft eine Variable von Interesse. Schon Kochmann (1923) bemerkt Unterschiede in der Behaltensleistung zwischen seinen Versuchspersonen, die er der unterschiedlichen musikalischen Begabung zuschreibt (Kochmann 1923, S. 337). Auch viele der modernen Studien zum Gedächtnis für kurze Tonfolgen integrieren die Dimension der musikalischen Expertise oder musikalischen Vorbildung als unabhängige Variable, obwohl diese meist nicht im Vordergrund der Untersuchungen steht. Gerade in einem varianzanalytischen Design scheint es aber sinnvoll, diese personenbezogene Variable einzubeziehen, um einen Teil der Varianz der Versuchsergebnisse erklären zu können und so für die eigentlich wichtigen Variablen (meistens musikalische Eigenschaften der Tonfolge) signifikante Ergebnisse zu erzielen. So kann beispielsweise die Hälfte der Varianz in experimentellen Daten von Monahan et al. (1987) durch die musikalische Expertise erklärt werden.

Die musikalische Expertise ist die einzige personenbezogene Variable, die regelmäßig in Untersuchungen zum Behalten von kurzen Tonfolgen betrachtet wird. Musikalische Expertise wird dabei fast ausschließlich über die Dauer oder den Grad der formalen musikalischen Ausbildung definiert. Andere personenbezogene Variablen wie die musikalischen Präferenzen, Intensität der Beschäftigung mit Musik, Gesangsfähigkeiten, Konzentrationsstärke, allgemeine Gedächtnisleistung, Alter oder Geschlecht wurden in der Vergangenheit so gut wie nicht erhoben, obwohl nicht im vorhinein ausgeschlossen werden kann, dass diese Variablen nicht im Zusammenhang mit der Behaltensleistung für kurze Tonfolgen stehen. Zwei Ausnahmen bilden die Untersuchung von Stacho & Nemeth (2001), die allgemeine Gedächtnisleistung und Gedächtnis für kurze Tonfolgen vergleicht, und die Studien von Mikumo (1992, 1994), die die Enkodierstrategie als unabhängige Variable aufnehmen.

Kauffman und Carlsen (1989) üben Kritik an der meist unzureichenden Differenzierung und der zu einfachen Definition der musikalischen Expertise über die Dauer der formalen musikalischen Ausbildung. Nicht nur sage die Dauer der Ausbildung nichts über deren Qualität und damit auch deren Erfolg aus, sondern die musikalische Expertise werde zusätzlich zum formalen musikalischen Training auch von einer Reihe anderer Faktoren bestimmt, die in der empirischen Literatur bis zum Ende der 80er Jahre nicht berücksichtigt worden seien (Kauffman & Carlsen, 1989, S. 4-5). Allerdings verwenden Kauffman und

Carlsen in derselben Studie eine Kombination aus Grad und Dauer der formalen musikalischen Ausbildung, um die Versuchspersonen entlang der Dimension der musikalischen Expertise einzuteilen. Auch sie verzichteten also auf eine differenziertere Erhebung der Dimension, die sich beispielsweise über das Erfragen und die gewichtete Verrechnung von mehreren Variablen erreichen ließe.

Als letzte wichtige Gruppe von unabhängigen Variablen, die in vielen Experimenten zum Behalten von kurzen Tonfolgen untersucht wurde, zählen die Variablen, die Varianten in den experimentellen Bedingungen selber beschreiben und nicht direkt mit der eigentlichen Konstruktion der Tonfolgen zusammenhängen. Diese Variablen sind oft an die experimentellen Bedingungen angelehnt, wie sie auch in Wiedererkennensparadigmen mit nicht musikalischen Stimuli betrachtet werden. Dazu zählen die Länge des Retentionsintervalls zwischen erster und zweiter Melodie (Dewar et al., 1977; DeWitt & Crowder, 1986; Dowling, 1991; Dowling et al., 1995; Stacho & Nemeth, 2001), die Präsentationsgeschwindigkeit der Töne (Dowling, 1972; Dowling, 1973; Edworthy, 1983, Tunks et al., 1993), die Position der Abweichung in der Tonfolge (Dewar et al., 1977; Cuddy & Lyons, 1981; Edworthy, 1983, Roberts et al., 1983; Roberts, 1986; Monahan et al., 1987; Cohen et al., 1989), die Anzahl der Töne (Long, 1977; Edworthy, 1983, 1985) und die Abfolge der *Testitems* innerhalb eines Versuchsdurchlaufes einer Versuchsperson (Dowling, 1973a; Eiting, 1984; Taylor & Pembroke, 1984). Wie die unabhängige Variable musikalische Expertise, sind auch diese Variablen in der Regel nicht der eigentliche Gegenstand des Erkenntnisinteresses, sondern dienen vor allem dazu, einen Teil der Varianz der Ergebnisse aufzuklären und so die Effekte der Manipulation der musikalischen Struktur signifikant werden zu lassen. In diesem Vorgehen liegt eine gewisse Plausibilität, wenn man bedenkt, dass die Versuchspersonen in vielen Fällen an sehr langen Experimentalsitzungen teilnehmen, die – wie schon oben angemerkt – durch die nicht nachvollziehbare Abfolge von sehr künstlichen Testitems geprägt sind. Es ist deshalb zu vermuten, dass die Resultate der Versuchspersonen zu einem nicht unerheblichen Teil durch die Anlage und die äußeren Bedingungen der Experimente beeinflusst sind und die Manipulationen der musikalischen Variablen nur für einen weiteren Teil der Ergebnisse verantwortlich sind.

3.2.5 Zusammenfassung der methodischen Schwierigkeiten der Experimente mit kurzen Tonfolgen

Die Idee eines wissenschaftlichen Experiments besteht darin, einen Ausschnitt der Welt nachzubilden und Erscheinungen zu beobachten, die durch die willkürliche Einwirkung des Beobachters entstanden sind. Aus diesen Beobachtungen werden Schlüsse auf Gesetzmäßigkeiten gezogen, die in diesem Ausschnitt der Welt wirken. Dabei ist selbstverständlich, dass es sich bei dieser Nachbildung der Welt nur um ein Modell und nicht um die Wirklichkeit selbst handelt, die sich notwendigerweise davon unterscheidet (vgl. Wundt, 1920 [1896]; Bredenkamp, 1996). Die Kritik an einem experimentellen Paradigma kann deshalb nicht darauf abzielen, einfach die Unterschiede zwischen Welt und Modell aufzuzählen, sondern muss darlegen, warum das Modell für das Studium des Ausschnittes der Welt nicht adäquat ist.

In diesem Sinne ist die folgende kurze Zusammenfassung der Kritikpunkte an den Forschungen mit kurzen Tonfolgen zu verstehen, die oben z.T. ausführlicher diskutiert wurden:

Vernachlässigung personenbezogener Variablen: Mit Ausnahme der Variable *musikalische Expertise* werden in den Experimenten mit kurzen Tonfolgen so gut wie keine Variablen einbezogen, die Eigenschaften der Versuchspersonen repräsentieren, obwohl diese durchaus für die Behaltensleistung eine Rolle spielen könnten. Zu den Variablen, die möglicherweise einen Einfluss haben, gehören die musikalischen Präferenzen, die Intensität der Beschäftigung mit Musik, Aufmerksamkeits- und Konzentrationsfähigkeit, die allgemeine Gedächtnisleistung, die Enkodier- und Abrufstrategien und die Gesangsfähigkeiten.

Ökologische Validität der kurzen Tonfolgen: Wie mehrere Autoren bemerkt haben, besitzen kurze Tonfolgen, wie sie in den hier diskutierten Experimenten verwendet wurden, nicht die musikalische Qualität von echten Melodien oder Melodiefragmenten. Es fehlen in der Regel einige musikalische Dimensionen wie Metrum, Rhythmus, Timbre, während die Gestaltung in anderen Parametern eher experimentellen als musikalischen Gesetzmäßigkeiten folgt, die sich oft aus der systematischen Variation der Versuchsbedingungen in einem (mehrfaktoriellen) varianzanalytischen Design ableitet; zu diesen Parametern zählen die melodische Kontur, die einer melodischen Linie inhärente Harmonik und die Abfolge von Sprüngen und Schritten. Es ist deshalb in vielen Fällen zu hinterfragen, ob die Erkenntnisse, die hinsichtlich der Behaltensleistungen von kurzen Tonfolgen experimentell

ermittelt wurden, auch auf das Gedächtnis für echte Melodien übertragbar sind oder ob Stimuli mit echter musikalischer Qualität auf andere Weise verarbeitet werden.

Künstlichkeit der Versuchsaufgaben: Die Aufgabe der Versuchspersonen in den meisten Experimenten besteht darin, die Identität zweier Tonfolgen oder die Identität hinsichtlich einer bestimmten Eigenschaft der Tonfolgen zu bestimmen. Dies ist keine Aufgabe, wie sie bei der alltäglichen Wahrnehmung und Verarbeitung besonders häufig vorkommt. In vielen Experimenten werden die Versuchspersonen deshalb vor dem eigentlichen Versuch für die Aufgabe trainiert. Dies lässt die Vermutung entstehen, dass die gefundenen experimentellen Ergebnisse für die Versuchsaufgabe zwar gültig sind, für die Gedächtnismechanismen echter Musikverarbeitung jedoch nur geringe Relevanz besitzen.

Reduktion der Antwortkategorien: Die Angaben der Versuchspersonen zur Identität zweier kurzer Tonfolgen werden meist in zwei Kategorien (richtig / falsch), ggf. noch mit Angabe der Urteilssicherheit festgehalten. Diese Reduktion der Antworten auf zwei Kategorien lässt in der Regel keine differenzierte Analyse der Fehler zu, die die Versuchspersonen beim Behalten der Tonfolgen machen und gibt keine Auskunft darüber, was die Versuchspersonen wirklich von einer Melodie behalten haben (vgl. die Kritik von Hantz, 1984). Wie Sloboda und Parker (1985) und Cutietta und Booth (1996) jedoch gezeigt haben (s. die Zusammenfassung dieser Studien im nachfolgenden Unterkapitel), können gerade Gedächtnisfehler und die Analyse von ungefähren Wiedergaben gute Einsichten in die Gedächtnismechanismen für musikalische Inhalte vermitteln.

Reduktion der musikalischen Merkmale: Durch die weit verbreitete Verwendung von Maßen aus der *Signal Detection Theory* als abhängige Variable ist eine Beschränkung der betrachtbaren Merkmale einer Tonfolge gegeben. Da diese Maße die Daten aus mehreren Durchläufen einer Vpn. unter einer Bedingung aggregieren, können die Tonfolgen durch nicht mehr als 2-3 Faktoren mit 2-4 Faktorstufen beschrieben werden. Alle darüber hinausgehenden Eigenschaften der Tonfolgen können nicht zur Erklärung der Gedächtnisleistungen herangezogen, sondern müssen als ‚Rauschen‘ betrachtet werden.

Annahme der ‚Unfehlbarkeit‘ des Identitätsurteils: Es wird angenommen, dass die Versuchspersonen in dem beschriebenen experimentellen Paradigma eine gehörte Tonfolge mit einer im Gedächtnis gespeicherten vergleichen und bei Identität der beiden Tonfolgen diese Identität angeben. Fehler bei dieser Angabe werden einzig dem musikalischen Gedächtnis, nicht aber dem Urteilsprozess zugeschrieben. Dabei ist nicht im vorhinein sicher gestellt, dass dieser Prozess immer fehlerfrei funktioniert, dass also die Versuchspersonen

auch bei perfekter Repräsentation der Tonfolge im Gedächtnis zu dem gleichen Urteil über die Identität wie der Experimentator gelangen.

3.2.6 Ergebnisse der experimentellen Literatur mit kurzen Tonfolgen

3.2.6.1 Melodische Kontur

Dowlings und Fujitanis einflussreiche Studie (1971) stellte die Frage nach der Bedeutung der melodischen Kontur für das Melodiegedächtnis. Diese Frage schien und scheint plausibel, wenn man die in weiten Teilen der abendländischen Kunstmusik verwendete Technik der Transposition von melodischen Motiven auf eine andere Tonleiterstufe in Betracht zieht. Bei diesen Transpositionen bleibt in der Regel die Kontur erhalten, während sich die Intervalle der zu Grunde liegenden Tonleiter anpassen. Trotz dieser Veränderung erkennen (mit dieser Musik vertraute) Hörer in der Regel ein solches Motiv auch in transponierter Form wieder. Tatsächlich bestätigen Dowling & Fujitanis Ergebnisse (1971, Exp. 1) diese Art des Erkennens bzw. der Speicherung melodischer Parameter im Gedächtnis: Ihre Versuchspersonen schienen sich bei transponierten kurzen Tonfolgen nur noch an der melodischen Kontur zu orientieren, während sie bei untransponierten offenbar auch die im Gedächtnis gespeicherten absoluten Tonhöhen der einzelnen Töne zum Wiedererkennen verwenden konnten. Dowlings Folgestudie (1972) zur Transformation von musikalischen Motiven erbrachte ganz ähnliche Ergebnisse hinsichtlich der melodischen Kontur. Und auch Massaro et al. (1980) konnten zeigen, dass unter verschiedenen Manipulationen von kurzen Tonfolgen diejenigen am wenigsten das Wiedererkennen beeinträchtigen, bei denen die Kontur erhalten bleibt. Dabei scheint eine intakte Kontur wichtiger zu sein als das Vorkommen von gemeinsamen Tönen.

Bestätigen konnten die Bedeutung der melodischen Kontur für das Wiedererkennen gerade bei kurzen Tonfolgen, die transponiert oder in einer anderen Form verändert werden, auch die Studien von White (1960), Dewar et al. (1977), Dowling (1978), Edworthy, (1983), Eiting (1984), Idson & Massaro (1976, 1978), Cutietta & Booth (1996).

Allerdings differenzieren diese in der Folge publizierten Studien die Bedeutung der Kontur oder zeigen einen Zusammenhang zu anderen Variablen der musikalischen Struktur oder der experimentellen Präsentationsform auf. Kontur scheint ein Parameter zu sein, der bei der Rekonstruktion der anderen Parameter einer kurzen Tonfolge aus dem Gedächtnis hilft, wie Cutietta & Booth (1996) schreiben.

In diesem Sinne ist die Kontur einer der beiden Bestandteile der Zwei-Komponenten-Theorie des Melodiegedächtnisse von Dowling (1978). Nach dieser Theorie kodiert der Hörer von einer kurzen Tonfolge die zu Grunde liegende Skala und die Kontur, die in dieser Skala ‚aufgehängt‘ wird. Kontur ist nach Dowling (1986) eine Art erster Index, der perzeptuellen Zugang zu genaueren Repräsentationen von beispielsweise Intervallen und Skalenstufen ermöglicht.

Ähnlich sieht Mindert Eiting (1984) die Kontur als erste Selektionsstufe einer Kaskade von Prozessen der Kodierung von kurzen Tonfolgen im Gedächtnis. Die Kodierung einer bestimmten Kontur beeinflusst nach Eiting die Verarbeitung der kurzen Tonfolge auf der nächsten Selektionsstufe, der relativen Intervallgröße. Am Ende steht nach dieser Theorie die Kodierung der exakten Intervallgröße innerhalb der ausgewählten Muster von Kontur und relativer Intervallgröße.

Auch Idson und Massaro (1978) fanden bei ihren Experimenten zum Wiedererkennen von Anfängen bekannter Melodien, dass die Repräsentation von Tonchroma nur dann zum Wiedererkennen benutzt werden kann, wenn gleichzeitig die Kontur der Tonfolge verwendet werden kann. Ähnlich bestätigte dies Dowling in späteren Experimenten (1984, 1986), mit denen er zeigen konnte, dass das Tonchroma allein für das Wiedererkennen einer Tonsequenz nicht ausreicht, sondern die melodische Kontur hierfür unbedingt erforderlich ist.

Einschränkungen für die Bedeutung der Kontur fand Judy Edworthy (1983, 1985). Offenbar ist nach ihren Ergebnissen die Kodierung der Kontur nur bei der Speicherung von kurzen Tonfolgen (3 bis 11 Töne) effizient. Bei längeren Tonfolgen (15 Töne) scheint die Tonhöheninformation bevorzugt gespeichert zu werden. Edworthy nimmt an, dass die Verarbeitung und Kodierung von Melodien hinsichtlich Kontur- und Tonhöheninformationen entlang eines Kontinuums verläuft und abhängig von der Länge der Melodie ist. Bei längeren Melodien bestünde für den Hörer die Möglichkeit, die Tonalität der Melodie zu extrahieren und Tonhöheninformationen im Bezug auf diese Tonalität zu speichern, was präziser und verlässlicher sei als die reine Konturinformation. Auch Dowling und Mitarbeiter (Dowling et al., 1995) fanden bei Tonfolgen, die im Rahmen von ‚echter‘ Musik präsentiert wurden und so in einem deutlich erkennbaren tonalen Kontext standen, dass die Konturinformation deutlich weniger für das Wiedererkennen benutzt wurde als dies ohne tonalen Kontext der Fall war.

Nicht nur die Länge der Tonfolge scheint mit der Bedeutung der Kontur für das Melodiegedächtnis zu interagieren, sondern auch die Länge des Retentionsintervalls, wie Dewitt

und Crowder (1986) und in der Tendenz auch Dowling und Bartlett (1981) fanden. Danach wird die Konturinformation eher dann zum Vergleich von zwei Tonfolgen im Gedächtnis benutzt, wenn diese rasch aufeinanderfolgen (1 Sekunde) als wenn das Retentionsintervall länger ist (25 bzw. 30 Sekunden). Bei längeren Retentionsintervallen scheinen Intervallinformationen wichtiger oder prominenter im Gedächtnis zu werden. Aus diesen Intervallinformationen werden – wie Dewitt und Crowder vermuten – bei längeren Retentionsintervallen Informationen zur Modalität oder Tonalität einer Tonfolge abstrahiert. Dies spricht für eine aktive Verarbeitung der kurzen Tonfolgen, während diese im Gedächtnis behalten werden. Diese Art von aktiver Verarbeitung nehmen auch ganz explizit Cutietta & Booth (1996, S. 234) für längere Zeitintervalle (Tage und Wochen) an. In dieser Studie wurden gesungene Wiedergaben von kurzen diatonischen Tonfolgen aufgezeichnet, die auf Fehler und Kategorisierungen der Tonfolgen durch die Versuchspersonen untersucht wurden. Neben der Kontur zeigte sich als wichtige Variable auch der Intervalltyp (Schritte vs. Sprünge) für die kategorische Einordnung und das Behalten der Tonfolgen.

Dass diese aktive Verarbeitung im Gedächtnis durch die Wahl bestimmter kognitiver Strategien zumindest bei musikalisch erfahrenen Vpn. noch unterstützt werden kann, zeigt die Studie von Mikumo (1994), in der die Vpn. durch ‚stummes Klavierspielen‘ die Tonsequenzen enkodierten und über lange Retentionsintervalle und störende Interferenzreize hinweg behalten konnten.

Dass die melodische Kontur allein für das Wiedererkennen von Tonfolgen mit gewisser musikalischer Qualität nicht ausreicht, stellte Dowling in einer Reihe von Experimenten Ende der 70er und Anfang der 80er Jahre fest, in denen die Tonalität von Test- und Vergleichssequenzen und deren Entfernung im Quintenzirkel manipuliert wurde. So fasst Dowling 1986 zusammen: „The relative importance of contour information in melody recognition varies with tonal scale context“ (Dowling & Harwood, 1986, S. 136). Der Grad, mit dem eine kurze Tonfolge die Wahrnehmung eines tonalen Zentrums evoziert und die Entfernung der tonalen Zentren in den beiden zu vergleichenden Tonfolgen sind also in hohem Maße mitentscheidend über die Art der Information, die eine Vpn. von einer solchen Sequenz im Gedächtnis behält und damit auch über die Strategie mit der die Vpn. den Gedächtnistest bewältigt. Diesen Einfluss der Nähe zu einem tonalen Zentrum und des Transpositions-niveaus auf die Wahrnehmung und Erinnerbarkeit von kurzen Tonfolgen konnten auch Cuddy et al. (1981) bestätigen; und auch Tunks et al. (1993) benutzen diese „tonal strength“ um die von ihnen im Experiment verwendeten Tonfolgen zu gliedern.

Taylor und Pembroke konnten in ihrer Studie (1984) noch genauer fassen, welche Art von Konturen das Behalten erleichtern bzw. den Parameter Kontur für Gedächtnisleistungen überhaupt erst nutzbar machen. In ihrer Studie, die in der Anlage eine Replikation der frühen Studie von Ortmann (1933) darstellt, fanden sie, dass Tonfolgen mit keinem oder nur einem Konturwechsel sowie mit einer symmetrisch angelegten Kontur sehr viel besser behalten werden als Tonfolgen mit mehreren Konturwechseln oder einer Kontur, die keine symmetrische Anlage aufweist. Diese bessere Verarbeitung und das daraus resultierende überlegene Wiedererkennen von sehr einfachen Konturen fanden auch schon Idson und Massaro in einer Serie von Experimenten, deren primäres Erkenntnisinteresse eigentlich Maskierungseffekten und dem Mechanismus der Streamsegregation galt (1976, Exp. I, II, IV).

Schließlich relativiert Dowling die Bedeutung der Kontur für das Gedächtnis für Tonsequenzen hinsichtlich der musikalischen Vorbildung (Expertise) der Versuchspersonen. Nach den Ergebnissen seiner Experimentalreihe von 1986 sind nur die Gedächtnisleistungen von Versuchspersonen mit mittlerer musikalischer Vorbildung mit seiner *scale-and-contour*-Theorie zu beschreiben. Anfänger orientierten sich bei diesen Experimenten eher an den Abständen der aufeinanderfolgenden Töne (Intervalle), Musiker schienen dagegen mehrere Informationsquellen zur Bewältigung der jeweiligen Aufgabe zu verwenden und miteinander zu kombinieren. Diese Flexibilität in den Gedächtnisstrategien und die daraus resultierenden bessere Leistungen bei Musikern konnte auch Mikumo (1992) nachweisen.

Zusammenfassend lässt sich hinsichtlich der Bedeutung der melodischen Kontur für das Behalten von kurzen Tonfolgen festhalten, dass der Kontur in den frühen Studien zu diesem Bereich große Bedeutung zugemessen wurde. Kontur schien eine Art *first-level*-Information zu sein, die Versuchspersonen schnell aufnehmen, stabil im Gedächtnis behalten und für die Rekonstruktion anderer musikalischer Parameter verwenden. Jedoch lässt sich aus Studien der letzten ca. 15 Jahre die Erkenntnis entnehmen, dass melodische Kontur nur einer unter mehreren für das Behalten relevanten Parametern ist. Auch die Erinnerung an eine melodische Kontur scheint mit der Zeit und durch Interferenzen anderer musikalischer Eigenschaften im Gedächtnis wandelbar zu sein. Zusätzlich scheint ein wichtiger Punkt für die Erinnerbarkeit der melodischen Kontur deren Einfachheit gemessen an der Zahl der Konturwechsel und dem symmetrischen Aufbau der Kontur zu sein. Nur einfache Kontu-

ren werden offenbar wirklich gut erinnert (z.B. Idson & Massaro, 1976; Taylor & Pembrook, 1984; Boltz & Jones, 1986).

Zumindest in einigen Studien scheint die geringere Bedeutung der Kontur durch die verwendeten Stimuli erklärbar zu sein. So verwendeten beispielsweise Dowling und Fujitani (1971, Exp. 1), Dowling (1972) und Massaro et al. (1980) sehr künstliche Stimuli (teilweise chromatische oder atonale Folgen von wenigen Tönen mit Sägezahn- oder Sinuswellenform, ohne Rhythmus und mit z.T. sehr kurzen Tondauern), die eine sehr geringe musikalische Qualität aufweisen (ähnlich auch Dowling, 1978). Taylor und Pembrook (1984) dagegen benutzten einen Piano-Sound für ihre durchweg diatonischen Tonfolgen, die auch Dreiklänge und melodische Sprünge enthielten. Ähnlich musikalisch gebaut waren auch die Testmelodien von Cutietta und Booth (1996), die zudem noch in gängigen Rhythmen und mit einem definierten Metrum zu hören waren. Es ist deshalb nicht unplausibel anzunehmen, dass vor allem bei Tonfolgen mit geringer musikalischer Qualität die melodische Kontur behalten wird. Je mehr Eigenschaften die verwendeten experimentellen Stimuli mit echten Melodien teilen, desto mehr dieser musikalischen Parameter können dann offenbar im Gedächtnis behalten werden, so dass sich die Bedeutung der Kontur relativiert und in der Hörwahrnehmung in einem Gefüge mit anderen musikalischen Parametern zusammenspielt.

Die Betonung der melodischen Kontur als für das Gedächtnis wichtigen Parameter steht mit Sicherheit auch in Verbindung mit dem von Dowling und anderen sehr häufig verwendeten experimentellen Design, in dem zwei transponierte Tonfolgen von den Vpn. direkt miteinander verglichen werden. Wie oben beschrieben ist die Kontur nur in diesem Falle von herausragender Bedeutung. Folgt dagegen die Testtonfolge untransponiert auf die Standardtonfolge, so werden auch die Informationsquellen absolute Tonhöhe und Intervallgröße sowie Skalenstufen des tonalen Kontextes zur Speicherung und Erinnerung verwendet.

Möglicherweise zeigt sich hier am Beispiel der melodischen Kontur, wie sehr Untersuchungsergebnisse, die für einen großen Bereich Gültigkeit beanspruchen, vom experimentellen Paradigma und den Versuchsstimuli abhängen. Diese Erkenntnis deuten auch Dowling und Harwood (1986, S. 152) in ihrer Zusammenfassung vor allem von Dowlings eigenen Studien selbst an: „We argue that task demands affect which melodic features are important in attending to and remembering melodies.“

3.2.6.2 Melodische Struktur und regelgeleiteter Aufbau

Seit den 70er Jahren rückte die Struktur und der regelhafte Aufbau von Musikstücken und Melodien in das Zentrum wissenschaftlicher Forschung. Den theoretischen Rahmen für diese Forschungen bildete zum einen die Informationstheorie und ihre Anwendungen in der Musikforschung (z.B. Moles, 1968; Berlyne, 1971, 1974) und die Diskussionen um Generationsgrammatiken (Sundberg & Lindblom, 1976; Deutsch & Feroe, 1981; Lerdahl & Jackendoff, 1983; Stoffer, 1985; s.a. Bruhn, 1988), die als Modelle bereits in der Sprachforschung erfolgreich gewesen waren.

Auch für das musikalische Gedächtnis wurde eine Speicherung von melodischen Informationen in abstrakten Strukturen bzw. Regeln oder in Abhängigkeit von informationstheoretischen Variablen angenommen.

So bemühen sich z.B. Cohen et al. (1989) in der Diskussion ihrer Experimente, Erinnerungs- und Identifizierungsleistungen bei melodischen Sequenzen auf die informationstheoretischen Parameter Redundanz und Unsicherheit zurückzuführen. Die Ergebnisse dieser Experimente lassen sich dahingehend zusammenfassen, dass eine geringere Unsicherheit beim Hörer bezüglich des Fortganges von zwei Vergleichssequenzen ihn in die Lage versetzt, Abweichungen leichter zu erkennen. Unsicherheit wird dabei über die Regelmäßigkeit und Einheitlichkeit des Aufbaus einer Sequenz operationalisiert. Obwohl Cohen et al. am Ende feststellen (Cohen et al., 1989, S. 27), dass dieser informationstheoretische Ansatz für die Untersuchung der Verarbeitung und Speicherung von melodischen Sequenzen ein sinnvoller Forschungsansatz ist, bleiben sie einen Hinweis schuldig, wie dieser Ansatz für Melodien oder ‚echte‘ Musik genutzt und überprüft werden kann.

Auch Cuddy et al. (1981) geben keinen direkten Nachweis, dass die von ihnen gefundenen Zusammenhänge zwischen regelhaftem Aufbau ihrer 7-Ton-Folgen und den Leistungen beim Entdecken von Unterschieden in diesen Tonfolgen auch für ‚echte‘ Musik gelten. Dennoch können sie mit Hilfe der Ergebnisse eines *Expertenratings* spezifizieren, welche Faktoren für einen regelhaften Aufbau ausschlaggebend sind: die Beziehung der Melodietöne zu einer (einfachen) harmonischen Struktur, die Einfachheit der melodischen Kontur und die Begrenztheit des melodischen Ambitus.

Mit ähnlicher Methodik gelangen Cuddy und Lyons (1981) zu einem sehr ähnlichen Ergebnis hinsichtlich der drei genannten Faktoren. Als Interpretation schlagen sie einen kognitiven Mechanismus bei der Verarbeitung bzw. Speicherung von melodischen Sequenzen unter Echtzeit-Bedingungen vor: Die Verarbeitung wird geleitet von der Tendenz, ein to-

nales Zentrum zu identifizieren. Dabei wird getestet, ob sich die aufeinanderfolgenden Töne durch einfache Transformationsregeln auf dieses tonale Zentrum beziehen lassen. Zu den einfachen Transformationsregeln gehören die Tonwiederholung, die Bewegung zum oder vom Grundton weg sowie das Einpassen in ein harmonisches Gerüst. Komplexe Konturen, große Sprünge und ungewöhnliche Harmonieschemata erschweren jedoch die Anwendung dieser Regeln in Echtzeitbedingungen und erklären so geringere Identifikations- bzw. Gedächtnisleistungen. Allerdings bemerken Cuddy und Lyons an den Strukturiertheit-*ratings* ihrer Experten, dass bei der Verarbeitung ohne Zeitbeschränkung von mehrdeutigen Tonfolgen, die Bewertungen und somit auch die Verarbeitung zwischen den Individuen sehr unterschiedlich verlaufen.

Eine formale Beschreibung der verwendeten Versuchsstimuli liefert Deutsch (1980). Dort werden melodische Sequenzen als Abfolge von hierarchischen Subsequenzen dargestellt, die sich als Operationen auf einem Tonalphabet in Bezug auf ein Referenzelement formulieren lassen. Deutsch greift dabei auf das von ihr zusammen mit Feroe entwickelte hierarchisch-generative Repräsentationssystem zurück (Deutsch & Feroe, 1981). Danach sind melodische Sequenzen einfacher und strukturierter, wenn sie sich sparsamer in diesem System kodieren lassen. Tatsächlich findet Deutsch (1980) eine fehlerfreiere Wiedergabe der einfacher strukturierten Testsequenzen. Für den kognitiven Speicherprozess nimmt sie deshalb an, dass die hierarchische Struktur und das Alphabet einer Sequenz vom Hörer erkannt und auch im Gedächtnis gespeichert wird. Zusätzlich muss dann nur noch ein einzelner *chunk*, z.B. der Anfang der Sequenz, mit der dazugehörigen Produktionsregel gespeichert werden, um die Tonfolge später aus dem Gedächtnis rekonstruieren zu können. Diese Art der Kodierung sei bezüglich der verbrauchten Gedächtniskapazität sehr effizient (Deutsch, 1980, S. 387). Darüber hinaus kann Deutsch in derselben Studie zeigen, dass die Alphabete (zugrundeliegende Tonleiter oder Tonalität) sehr gut auch unabhängig vom konkreten melodischen Verlauf erinnert werden. Dieses Ergebnis wurde in ähnlicher Weise schon von Dowling (1978) vorweggenommen. Schließlich beschreibt Deutsch, dass für das Erkennen der melodischen Struktur durch die Versuchspersonen die zeitliche Segmentierung (Position der Pausen zwischen Tönen) eine entscheidende Rolle spielte. Die Sequenzen, bei denen die zeitliche und melodische Segmentierung zusammenfiel, wurden besser erinnert als die unterschiedlich segmentierten. Deutsch interpretiert diese Ergebnisse dahingehend, dass die Versuchspersonen melodische Sequenzen gemäß ihrer hierarchischen Struktur und den diskutierten Produktionsregeln verarbeiten und erinnern. Einschränkung-

gen erfahren diese Ergebnisse jedoch einerseits durch die Tatsache, dass ihre 10 bzw. 12 Versuchspersonen ausschließlich aus musikalisch gut trainierten Studenten rekrutiert waren, denen das Denken in musikalischen Produktionsregeln mit Sicherheit nicht so fern liegt wie Personen ohne musikalische Vorbildung. Andererseits gibt Deutsch, wie auch Cohen et al. und Cuddy (s.o.), keinen Hinweis darauf, wie die Variablen Strukturiertheit und Einfachheit sowie Übereinstimmung von melodischer und zeitlicher Segmentierung in ‚echter‘ Musik bestimmt und quantifiziert werden können. Insofern ist das Testen ihrer Voraussagen an echter Musik nicht ohne weiteres möglich.

Die Interpretation, dass Versuchspersonen Transformationen von melodischen Tonfolgen (unbewusst) verarbeiten und abstrahieren können, stützen einige Experimente, die die Abstraktion von Prototypen aus melodischen Varianten (Welker, 1982; Carterette et al. 1986; Bigand, 1990; Kötter, 1991; Kötter 1992b) untersuchen.

Ähnlich wie Deutsch (1980) kommen mit dieser Technik auch Carterette et al. (1986) zu dem Ergebnis, dass die Regeln von Transformationen, wie Veränderungen der Tondauer, der Tonhöhe und der Kontur, von den Versuchspersonen kognitiv nachvollzogen und verarbeitet werden können. Dabei sind die Einfachheit bzw. die Anzahl der Transformationen, denen eine Tonsequenz unterworfen wird von entscheidender Bedeutung: Tonsequenzen, die nur eine Transformationen ‚auseinanderliegen‘, werden als kognitiv ähnlicher beurteilt als solche, die sich durch Anwendung von zwei Transformationen unterscheiden. Die Versuchspersonen können also offensichtlich die objektive Anwendung von Transformationen subjektiv nachvollziehen. Als Nachweis nehmen Carterette et al. sowie Welker (1982) auch das fälschliche Wiedererkennen des Prototypen, der Ausgangspunkt für die nach bestimmten Regeln transformierten Tonfolgen gewesen war.

Bigand (1990) führt die Ergebnisse seines Experimentes auf das Erkennen der zugrundeliegenden harmonischen Struktur durch die Versuchspersonen zurück. Wie Bigand in Anlehnung an Lerdahl und Jackendoff (1983) vermutet, lässt sich eine konkrete melodische Folge durch gewisse Regeln aus einer harmonischen Struktur generieren. Umgekehrt ist es für den Hörer möglich, durch Rückverfolgen der Transformationen von der Tonfolge auf die Harmoniefolge zu schließen.

Kötter knüpft in den mehrfach veröffentlichten Ergebnissen seines Experimentes (Kötter, 1991, 1992a, 1992b) an die Versuche von Welker und Bigand an und bestätigt den kognitiven Mechanismus der Abstraktion von Prototypen aus melodischen Varianten, wobei er echte melodische Varianten eines Themas aus Mahlers 3. Sinfonie benutzt. Jedoch lehnt

Kötter das Aufsummieren und Kombinieren von regelgeleiteten Transformationen als Erklärung der Abstraktionsbildung ab. Nach seiner eher vage gehaltenen Darstellung der Prototypenbildung erfolgt diese „durch ganzheitliche von Vorerfahrung geprägte Wahrnehmung“, bei dem neben der Harmonik auch die von Welker und Bigand nicht beachtete rhythmische Komponente eine Rolle spielt (Kötter, 1992a, S. 145).

Große Zweifel an der Bedeutung von rekursiven Generations- und Transformationsregeln für die kognitive Verarbeitung und das musikalische Gedächtnis äußern Boltz und Jones (1986). Die Experimente dieser Studie gehen der beispielsweise von Deutsch (1980) oder Simon und Sumner (1968) aufgestellten Hypothese nach, dass Tonfolgen, die rekursiv durch iterative Regelanwendung definiert und deshalb formal einfacher beschrieben werden können, leichter zu behalten und zu reproduzieren seien. Gemäß dieser Hypothese soll dem einfacheren formalen Code ein einfacherer mentaler Code entsprechen, den die Versuchspersonen erzeugen und so die Erinnerungsleistung in der kapazitätsbegrenzten Resource Gedächtnis erhöhen können. Boltz und Jones fanden jedoch, dass die rekursive Anwendung von Regeln bei der Generierung von kurzen Tonfolgen nicht deren Erinnerbarkeit erhöht. Statt dessen scheinen eher die Einfachheit der melodischen Kontur und die geringere Zahl bzw. der regelmäßigere Einsatz von Intervallsprüngen sowie die zeitliche Regelmäßigkeit dieser melodischen Ereignisse die Erinnerbarkeit einer Tonfolge zu erhöhen (Boltz & Jones, 1986, S. 405). Boltz und Jones entwickeln aus diesen Erkenntnissen ein dynamisches Akzentmodell, welches die leichtere Verarbeitung und Erinnerbarkeit von Tonfolgen aufgrund wiederkehrender melodischer und rhythmischer Akzente und des Zusammenhangs dieser beiden Akzentarten voraussagt (Boltz & Jones, 1986, Exp. 2; Jones et al., 1987, Jones & Boltz, 1989; Boltz, 1991, 1993).

Ein ganz ähnliches Modell, welches die strukturell wichtigen Töne und damit die Erinnerbarkeit von Tonsequenzen bzw. Melodien aus dem Zusammentreffen von rhythmischen und Tonhöhenakzenten vorhersagt, entwerfen Monahan et al. (1987). Hier spielen Erwartungen über den Fortgang einer Melodie eine große Rolle, die aus der musikalischen Erfahrung der Vpn. aufgrund des Beginns eines konkreten Melodie konstruiert werden. Diese Art von Erwartungen sind auch das zentrale Konzept in einer Reihe von Modellen anderer Autoren, die ab Ende der 80er Jahre entstanden (z.B. Schmuckler, 1989; Smith & Cuddy, 1989; DeWitt & Samuel, 1990). Da diese Modelle sowohl von der Vorerfahrung ausgehen, die Versuchspersonen anhand von Melodien gewonnen haben und gleichzeitig realistischere Stimulusmaterialien oder sogar ‚echte‘ Musik im Experiment einsetzen, soll dieser

Ansatz im folgenden Unterkapitel 3.3 „Gedächtnis für Melodien“ ausführlich vorgestellt und diskutiert werden.

Es lässt sich also zusammenfassen, dass wie für den gesamten Bereich der kognitiven Musikwissenschaft in den 70er und 80er Jahren auch für viele Forschungen zum Gedächtnis für kurze Tonfolgen Ansätze aus der Informationstheorie sowie aus der Linguistik entlehnte Modelle von Generationsgrammatiken sehr einflussreich waren. In vielen Studien erwiesen sich die untersuchten Parameter, wie Unsicherheit, formale Einfachheit der Produktionsregeln oder Anzahl ausgeführter regelgeleiteter Transformationen als tatsächlich signifikante Einflußgrößen für die Erinnerungsleistung.

Manko all dieser Forschungen ist jedoch, dass sie einzig mit künstlich generierten Tonfolgen als Stimulusmaterial arbeiten. Insofern bleibt ungeklärt, wie diese Einflussgrößen in echten Melodien bestimmt und quantifiziert werden können und ob sie bei der Erinnerung von Melodien aus echter Musik noch den Einfluss besitzen, den sie in Experimenten mit künstlich generierten Stimuli zeigten.

Kötter (1991; 1992b) weist auf dieses Problem hin und verwendet für sein eigenes Experiment Fragmente aus echten Melodien. Jedoch bleibt er den Vorschlag eines konkreten, generalisierbaren und überprüfbaren Modells schuldig.

In ihrem zusammenfassenden Überblick geht Christa Nauck-Börner (1988) noch auf einige weitere Probleme dieser Struktur- oder Grammatikmodelle zum Melodiegedächtnis ein. Zu den Einschränkungen bzw. Problemen dieser Modelle, die Nauck-Börner aufzeigt, gehören die oft ungeprüfte Übernahme von musikanalytischen Regeln und Techniken in psychologische Modelle, die Nicht-Beachtung von psychischen Kapazitätsgrenzen (z.B. Gedächtniskapazitäten in Abhängigkeit von musikalischer Vorerfahrung) und die primäre Orientierung an Intervallstrukturen ohne Berücksichtigung der Melodiekontur. Als Perspektive für effektivere und realitätsnähere Modelle zum Melodiegedächtnis spricht sich Nauck-Börner für Schemamodelle aus, innerhalb derer stilistische musikalische Eigenheiten und Vorwissen von Hörern mitbeachtet werden können. Sie selbst entwirft in ihrem kurzen Artikel jedoch kein Beispielmmodell, das den Schemaansatz exemplifizieren würde.

Einen Schritt über die Generationsgrammatiken hinaus und vielleicht schon in Richtung eines Schemaansatzes gehen die Akzentmodelle von Boltz und Jones bzw. Monahan und Mitarbeitern (Boltz & Jones, 1986; Jones & Boltz, 1989; Monahan et al., 1987). Diese

Modelle wurden zunächst zwar an künstlich generierten kurzen Tonfolgen entwickelt und getestet, lassen sich aber für echte Melodien generalisieren; d.h. es werden überprüfbare Voraussagen über die Erinnerbarkeit von echten Melodien aufgrund ihrer strukturellen Eigenschaften möglich. Aus diesem Grund sollen diese vielversprechenden Modelle im Unterkapitel 3.3 „Gedächtnis für Melodien“ erörtert werden.

3.2.6.3 Zusammenfassung der Erkenntnisse zum Gedächtnis für kurze Tonfolgen

Kurze Tonfolgen waren in den letzten 30 Jahren als Stimulusmaterial für sehr viele Experimente zum musikalischen Gedächtnis von größter Wichtigkeit. Mit ihrer Hilfe schien es sehr gut möglich, die für die Behaltensleistung relevanten Bedingungen und Faktoren der musikalischen Struktur gemäß der experimentellen Zielsetzung zu erzeugen und nach der varianzanalytischen Logik systematisch zu variieren.

Wie jedoch einige Forscher bemerkt haben, gibt es Grenzen für die Übertragbarkeit der mit kurzen Tonfolgen ermittelten experimentellen Ergebnisse auf den Bereich des Gedächtnisses für echte Melodien. Besondere Kritik gilt dabei vor allem der Künstlichkeit der experimentellen Aufgaben und der kurzen Tonfolgen selbst, die sie für die meisten Versuchspersonen als wenig musikalisch erscheinen lassen dürfte und deshalb unter Umständen eine Art der Verarbeitung und Speicherung provoziert, die nicht dem kognitiven Umgang mit echten Melodien entspricht. Zusätzlich versäumen es viele der in diesem Teilkapitel diskutierten Studien, das Gedächtnisverhalten, das die Versuchspersonen mit den kurzen Tonfolgen zeigen, umfassend zu erheben und auszuwerten. In der Regel werden die Antworten der Versuchspersonen auf die Kategorien ‚richtig/falsch‘ reduziert und auf eine Analyse der Gedächtnisfehler ebenso verzichtet wie auf die Untersuchung personenbezogener Variablen, die im Zusammenhang mit dem Gedächtnis für kurze Tonfolgen stehen könnten (z.B. allg. Gedächtnisleistung und Konzentrationsfähigkeit, Gesangsfähigkeit oder Intensität der Musikbeschäftigung).

Diese methodischen und erkenntnistheoretischen Probleme machen es neben den vielen divergenten Interpretationsansätzen und Ergebnissen sehr schwierig, einige gesicherte Erkenntnisse zu den Gedächtnisleistungen für kurze Tonfolgen oder gar kurzen Melodien ‚herauszudestillieren‘. Die Ergebnistabelle zum Schluss dieses Kapitels ist deshalb unter dem Vorbehalt zu lesen, dass die präsentierten Ergebnisse strenggenommen nur Gültigkeit im Rahmen der jeweiligen Versuchsanordnung haben und auf den Bereich des Gedächtnisses für echte Melodien nur beschränkt übertragbar sind. Zusätzlich lassen sich diese Er-

gebnisse nicht im Rahmen eines einzigen Gedächtnismodells interpretieren, sondern sind vielmehr eine Sammlung von Einzelergebnissen, die unverbunden nebeneinander stehen und auf verschiedene, wahrscheinlich konkurrierende Gedächtnis- und Verarbeitungsprozesse zurückgehen. Um jedoch einen ersten Überblick über die Forschungsansätze und die Qualität der Ergebnisse zum Gedächtnis für kurze Tonfolgen bekommen, dürfte die Tabellendarstellung hilfreich sein.

Die Vielzahl der in diesem Kapitel und in der abschließenden Tabelle präsentierten Ergebnisse lässt sich trotz der vorgenannten Einschränkungen mit einem Erkenntnisgewinn anhand der musikalischen Qualität der verwendeten Stimulusmaterialien bzw. ihrer Nähe zu echten Melodien zusammenfassen.

Wie die Diskussion um die Bedeutung der melodischen Kontur für das Erinnern von kurzen Tonsequenzen zeigte, ist Kontur besonders unter drei Bedingungen für das Behalten effektiv: Zum ersten, wenn sie die einzige sinnvolle oder nachvollziehbare Dimension der präsentierten Tonfolgen ist. Zum zweiten, wenn sie in Interaktion mit bestimmten anderen musikalischen Parametern wie der zugrundeliegenden Skala oder der Tonalität wahrgenommen wird und so eine Rekonstruktion anderer Parameter wie etwa der Intervallstruktur ermöglicht. Zum dritten ist die melodische Kontur dann eine für das Gedächtnis bedeutende Dimension, wenn sie sich in den präsentierten Stimuli als besonders einfach oder symmetrisch darstellt, wie es in vielen prägnanten melodischen Floskeln in ‚echter‘ Musik gegeben ist. Die Bedeutung der melodischen Kontur für das Behalten scheint also davon abzuhängen, wie musikalisch oder musikähnlich eine bestimmte Stimulusfolge insgesamt gebaut ist.

Ein ähnliches Ergebnis lässt sich auch aus den Ergebnissen der Studien zum Einfluß der Struktur und des regelhaften Aufbaus von Tonfolgen entnehmen. Betrachtet man die Ergebnisse von Deutsch (1980) oder von Cohen et al. (1989), so ist auffällig, dass gerade diejenigen Testsequenzen besser erinnert werden, die echten Melodien oder Melodiefloskeln am ähnlichsten sind (z.B. geläufige Intervalle und regelmäßige Wiederholungen). Die Behandlung dieser Testsequenzen nach den Mechanismen der Verarbeitung echter Melodien wird dabei vermutlich auf mehreren Ebenen (harmonisch, rhythmisch, Assoziation zu bekannten melodischen Floskeln etc.) möglich. Cuddy (Cuddy & Lyons, 1981; Cuddy et al. 1981) liefert sogar innerhalb ihres informationstheoretischen Ansatzes ein Anforderungsprofil für Tonsequenzen, die gut erinnert werden können, welches quasi als Beschreibung von musikalischer Qualität oder ‚Musikähnlichkeit‘ gelesen werden kann:

Die Einzeltöne müssen eine Beziehung zu einer einfachen harmonischen Struktur haben – dies schränkt auch die Wahl der Intervalle ein –, die melodische Kontur und die verwendeten Intervalle müssen einfach bzw. geläufig sein, der melodischen Ambitus darf nicht zu groß sein.

Ganz ähnlich lassen sich auch die Ergebnisse der Studien interpretieren, die die Abstraktion von Regeln und Prototypen aus einer Menge melodischer Varianten untersucht haben. Insgesamt lässt diese Lesart jenseits der konkreten theorieverbundenen Interpretationsansätze den Schluß zu, dass eine kurze Tonsequenz desto besser erinnert werden kann, je mehr Eigenschaften ‚echter‘ Melodien sie aufweist. Die Bedeutung der einzelnen strukturellen Dimensionen für die Erinnerung scheint dabei vor allem vom Bau und der konkreten Gestalt der Sequenz selbst abzuhängen. So wird beispielsweise von einer kurzen Tonfolge ohne Rhythmus und mit ungewöhnlichen Intervallen, die das Erschließen zugrundeliegender Harmonien oder einer Tonalität erschweren, vor allem die Kontur erinnert werden, da sie noch am ehesten mit realen Konturen vergleichbar ist.

Zu einem sehr ähnlichen Ergebnis gelangte bereits Kochmann (1923), der feststellte, dass seine Versuchspersonen von den Tonsequenzen, die ein Tonalitätsgefühl aufkommen ließen bzw. einen „geschlossenen Charakter“ aufwiesen, weit mehr Töne aus dem Gedächtnis richtig wiedergeben konnten als von den Sequenzen, bei denen diese Eigenschaften nicht gegeben waren.

Einen deutlichen Hinweis darauf, dass die Versuchspersonen bei der Verarbeitung und Speicherung eine Verbindung zu echten Melodien herstellen, lieferte schließlich auch die Fehleranalyse, der Cutietta und Booth (1996) die Wiedergaben von kurzen diatonischen Tonfolgen aus dem Gedächtnis unterzogen. So kombinierten die Versuchspersonen von Cutietta und Booth häufig zwei verschiedene vorher gelernte Melodien auf ‚musikalische Weise‘, wobei sie unter anderem oft den Modus (Dur statt Moll und umgekehrt) konsequent änderten, die Tonfolgen häufig abweichend vom Original mit dem Übergang von Leitton zu Grundton beendeten und besonders bei hohen Tönen viele Ungenauigkeiten in der Tonhöhe zeigten. Die Versuchspersonen ‚musikaliserten‘ also die kurzen Tonfolgen im Gedächtnis, ganz ähnlich wie es Kochmanns Versuchspersonen taten. Diese nivellierten beispielsweise die z.T. mehrdeutige Tonalität der kurzen Tonfolgen oder sie ersetzten bei Unsicherheiten einzelne Töne durch die quasi als *default-Werte* verwendeten Intervalle große Terz, Quinte oder große Sexte relativ zum wahrgenommenen Grundton (Kochmann, 1923, S. 348).

Vielversprechend in diesem Zusammenhang sind deshalb Ansätze, die von der Vorerfahrung von Versuchspersonen mit echten Melodien ausgehen und aufgrund dieser Vorerfahrungen und mit Hilfe definierter Regeln Voraussagen über die Erinnerbarkeit einer konkreten Test-Melodie treffen können, wie es die Modelle von Boltz und Jones oder Monahan und Mitarbeitern tun. Diese Modelle sollen im folgenden Kapitel über echte Melodien diskutiert werden.

3.2.7 Empirische Ergebnisse der Forschungen zum Gedächtnis für kurze Tonfolgen (Tabelle)

Ergebnis	Kommentar	Beleg
Anzahl Töne, Dauer Retentionsintervall		
Kürzere Tonfolgen (ca. 5 - 7 Töne) werden besser wiedererkannt als längere (z.B. 7 - 15 Töne)		Long (1977); Edworthy (1985); Dewitt & Crowder (1986)
Bei transponierten kürzeren Tonfolgen (z.B. 5 Töne) wird Kontur am besten verarbeitet, bei längeren (z.B. 15 Töne) Intervall- und Tonhöheninformation		Edworthy (1983); Edworthy (1985)
Tonfolgen werden nach kurzem R-Intervall (1 bzw. 7 Sek.) besser erkannt als nach langem (30 bzw. 33 Sek.)		Dewitt & Crowder (1986); Dowling et al. (1995, Exp. 4)
Nach kurzem R-Intervall wird Kontur am besten wiedererkannt, bei längerem R-Intervall Modalität/Intervallinformation		Dowling & Bartlett (1981); Dewitt & Crowder (1986); Dowling (1991); Dowling et al. (1995);
Präsentationsgeschwindigkeit (bpm)		
Wiedererkennen (Speichern) von kurzen Tonfolgen funktioniert nur zw. 100-240 bpm (bei Präsentation in 4teln) optimal; über und unter diesen Grenzen Leistungseinbrüche		Tunks et al. (1993)
Melodische Kontur vs. Intervallinformation		
Kontur wird v.a. dann als melodischer Parameter gespeichert und wiedererkannt, wenn Tonfolge transponiert ist oder andere Parameter nicht musikalisch sinnvoll oder zu schwach ausge-	Sind Ergebnisse bzw. Interpretationen der entsprechenden Studien;	Dowling (1978); Bartlett & Dowling (1980); Dowling & Bartlett (1981); Massaro et al (1980); Edworthy

prägt sind oder die Tonfolgen in keinem musikalischen Kontext stehen;		(1985); Dowling & Harwood (1986); Dowling et al. (1995); Cutietta & Booth (1996)
Bei nicht transponierten Vergleichsmelodien wird Intervallinformation genauso gut verarbeitet wie Konturinformation;	Sobald ein tonaler Rahmen kognitiv etabliert ist, kann Intervallinformation effektiv verarbeitet werden (Edworthys Interpretation)	Edworthy (1985);
Kontur muss einfach sein (wenige Richtungswechsel, symmetrische Anlage), damit sie gespeichert bzw. richtig wiedererkannt wird	Einfachheit von melodischer Kontur ist nicht exakt definiert	Idson & Massaro (1976); Cuddy & Lyons (1981); Taylor & Pembrook (1984); Boltz & Jones (1986)
Kontur hilft / ermöglicht Rekonstruktion anderer melodischer Parameter von Tonfolge im Gedächtnis;		Idson & Massaro (1978); Dowling (1978); Eiting (1984); Cutietta & Booth (1996);
Enkodierung von Kontur ist bewusster Prozess, Enkodierung von Intervallinformationen bzw. Skaleninformation ist automatischer Prozess;		Dowling et al. (1995);
Absolute Tonhöhe und Skala		
Absolute Tonhöhen der gemeinsamen Töne von zwei Tonfolgen werden zum Wiedererkennen verwendet, wenn beide Vergleichstonfolgen auf selbem Tonhöheniveau liegen.		Dowling & Fujitani (1971); Dowling et al. (1995);
Zugrundeliegende Skala bzw. Tonart wird aus kurzen Tonfolgen unabhängig vom konkreten melodischen Verlauf abstrahiert und gespeichert		Dowling (1978); Deutsch (1980); Dowling et al. (1995);
Abstraktion von Strukturen		
Aus konkreten Tonfolgen können Strukturen abstrahiert und gespeichert werden. Mögliche Strukturen sind harmonisches Schema, Transformationsregeln und die Wiederholung von Motiven und Konturschema.	Die Beschreibung der kognitiv abstrahierten Strukturen aus echten Melodien ist nicht geklärt	Deutsch (1980); Welker (1982); Carterette et al. (1986); Bigand (1990); Kötter, (1992a);
Je ähnlicher Strukturen denen echter Musik sind, desto besser werden sie gespeichert bzw. wiedererkannt	Interpretation der Ergebnisse der entsprechenden Studien	Kochmann (1923); Cuddy et al. (1981); Cuddy & Lyons (1981); Boltz & Jones (1986); Cutietta & Booth (1996);
Musikalisches Vorwissen		
Versuchspersonen mit (guter) musi-		Long (1977); Dowling

kalischer Vorbildung erkennen kurze Tonfolgen besser wieder als Vpn. ohne bzw. mit geringer musikalischer Vorbildung;		(1986); Monahan et al., (1987); Mikumo (1992); Dowling et al. (1995);
--	--	---

3.3 Gedächtnis für Melodien

Während die Forschungen, die kurze Tonfolgen als Stimulusmaterial für Gedächtnisexperimente verwenden, sich hinsichtlich Methodik und Erkenntnisinteresse als relativ homogener Forschungskorpus darstellen lassen (s. Kap. 3.2), werden in diesem Kapitel z.T. recht divergente Ansätze der Erforschung des Melodiegedächtnisses besprochen. Die kleinsten gemeinsamen Nenner der in diesem Kapitel diskutierten empirischen Untersuchungen sind zum einen das Melodiegedächtnis als zentraler Gegenstand der Forschungsinteressen und zum anderen die Verwendung von Tonfolgen mit meistens mehr als 15 Tönen als Stimulusmaterial. Fast ausnahmslos sind ‚echte‘ Melodien, wie sie in realen, nicht für das psychologische Experiment komponierten Musikstücken vorkommen, zwar die Inhalte, die hier von eigentlichem Interesse sind, jedoch werden zumeist aus Gründen der experimentellen Kontrolle eine ganze Reihe von verschiedenen Stimulusmaterialien in den einzelnen Experimenten verwendet. Dazu zählen neben Melodien aus existierenden Musikstücken, die den Versuchspersonen bekannt (z.B. Dowling et al., 1995; Hébert & Peretz, 1997) oder unbekannt (z.B. Sloboda & Parker, 1985; Oura & Hatano, 1988; Dowling et al., 1995; Gardiner et al. 1996) sind, auch neu komponierte Melodien (Bartlett et al., 1995), nach Regeln konstruierte längere Tonfolgen (Boltz & Jones, 1986; Monahan et al., 1987), ‚verunstaltete‘ und ‚maskierte‘ bekannte Melodien (Deutsch, 1972; Dowling & Hollombe, 1977) und melodische Passagen aus echten, z.T. mehrstimmigen Musikstücken (Russell, 1987; Kauffman & Carlsen, 1989; Schmuckler, 1990; Dowling et al., 2002). Der Begriff *Melodie* wird in diesem Kapitel also mehr im Sinne einer Abgrenzung zu den *kurzen Tonfolgen* mit meist sehr künstlichem Charakter des vorigen Kapitels verwendet. Zwar sind definitorisch in der Regel Melodien in der Menge der Tonfolgen eingeschlossen, die Verwendung des Terminus *Melodie* in diesem Unterkapitel soll jedoch ausdrücken, dass die gemeinten Testmaterialien mehr musikalische Qualitäten besitzen als die mit Tonfolgen bezeichneten Stimuli des vorangegangenen Teilkapitels. Dennoch erfüllen auch nicht alle Versuchstimuli der in diesem Kapitel besprochenen Studien die Ansprüche an Geschlossenheit oder melodische Qualität, wie sie ‚gute‘ Melodien aus existenten Liedern oder Musikstücken aufweisen.

Wegen der ganz verschiedenen Zielsetzungen und Experimentaldesigns der in diesem Kapitel zu diskutierenden Studien, ist es nicht sinnvoll, wie im vorangegangenen Kapitel Forschungsziele, Methodiken und Ergebnisse aller Untersuchungen zum Melodiegedächtnis

gemeinsam zu besprechen, sondern eine Gliederung nach den einzelnen Untersuchungsansätzen scheint für die Aufbereitung dieser Literatur geeigneter zu sein.

Die Zuordnung der Studien zu den verschiedenen Unterabschnitten ist dabei vor allem nach zwei Kriterien vorgenommen worden: Der Überblick über die Literatur soll einfach nachvollziehbar sein, und die für den Teil II dieser Arbeit nutzbaren Ergebnisse sollen miteinander verbunden dargestellt werden.

3.3.1 Wiedergabe unbekannter Melodien

Die im letzten Kapitel bereits angesprochene Studie von John Sloboda und David Parker (1985) eröffnete eine neue Perspektive für die Erforschung des Musikgedächtnisses. Nicht nur wurden das allgegenwärtige Wiedererkennensparadigma und die damit verbundenen Mess- bzw. Erhebungsverfahren sowie die Fixierung auf kurze Tonfolgen hinsichtlich der ökologischen Validität (s. Kap. 3.2) kritisiert, sondern Sloboda und Parker konnten auch durch die Verwendung eines anderen Paradigmas und anderer Stimuli im Experiment zu deutlich anderen Erkenntnissen über die Prozesse des Melodiegedächtnisses gelangen. Nicht die Bedeutung einzelner Dimensionen bei der Speicherung von melodischen Folgen, sondern die Verarbeitung einer ganzen Melodie mit ihren globalen Merkmalen wie formaler Bau, Abschnittssegmentierung, Metrum oder harmonisches Gerüst sind die wichtigen Ergebnissen dieser Studie.

Slobodas und Parkers Ansatz und Methode wurden in der Folge von mindestens zwei weiteren Studien weiterentwickelt und differenziert. Diese beiden Studien von Oura und Hatano (1988) und Zielinska und Miklaszewski (1992) sollen zusammen mit der von Sloboda und Parker (1985) kurz zusammengefasst werden.

Methode:

Ausgangspunkt für Sloboda und Parker (1985) war die Verwendung eines Wiedergabeparadigmas statt des bis dahin fast ausschließlich verwendeten Verfahrens der Wiedererkennung bzw. Identifizierung zum Überprüfen von melodischen Gedächtnisinhalten. Wie oben in den Abschnitten 3.2.3 und 3.2.4.2 ausführlich besprochen, besteht der Nachteil bei der Verwendung eines Wiedergabeparadigmas in Experimenten zum Melodiegedächtnis darin, dass die in der Regel gesungenen Wiedergaben erst transkribiert und dann mit nicht-standardisierten Verfahren aufbereitet und analysiert werden müssen. Die so erhobenen Daten nennen Sloboda und Parker „messy“ (S. 147), da die Zuordnung der gesungenen Wieder-

gaben zu numerischen Daten und die Konstruktion von geeigneten Messskalen im Vergleich mit den im Wiedererkennungsverfahren erhobenen Daten sehr aufwendig und mühsam ist. Dieser Nachteil wird jedoch durch die Erhebung von weit mehr verwertbaren Informationen inklusive der Möglichkeit zu einer detaillierten Analyse der Gedächtnisfehler und durch die deutlich größere ökologische Validität der gesamten Experimentalsituation kompensiert.

In Slobodas und Parkers Versuch wurde den Versuchspersonen einzeln jeweils sechsmal eine unbekannte russische Volksmelodie von 30 Tönen (6 Takte) mit einem echten Klaviersound vorgespielt. Nach jedem der sechs Durchläufe sollte die Vpn. die Melodie gesungen wiedergeben. Jede Wiedergabe wurde aufgenommen und anschließend transkribiert sowie auf eine einheitliche absolute Tonhöhe transponiert. Die Transkriptionen wurden schließlich unter verschiedenen Aspekten, wie Metrum, Atempausen, melodische Kontur, Phrasenstruktur und Unterschiede in Bezug auf die musikalische Vorbildung der Versuchspersonen analysiert. Dabei war Slobodas und Parkers Aufmerksamkeit weniger auf die Bedeutung bzw. Gewichtung der verschiedenen Parameter untereinander gerichtet, als auf die Entwicklung der Wiedergaben innerhalb dieser Dimensionen über die sechs Versuchsdurchgänge hinweg und auf die Art der Abweichungen von der Originalmelodie.

Ein stärker formalisiertes und daher für vergleichende Analysen besser geeignetes Auswertungsverfahren entwickelten Oura und Hatano (1988). Dieses Auswertungsverfahren wurde auch von Zielinska und Miklaszewski (1992) und in Abwandlungen auch von Drake et al. (1991) und Ogawa et al. (1995) zur Untersuchung von leicht abweichenden Fragestellungen (s.u.) verwendet. Das Auswertungsverfahren besteht im Wesentlichen aus schrittweisen Bewertungen, bei denen die auswertende Person jeweils halbe Takte von Originalmelodie und Wiedergabe vergleicht:

Test auf tonale Stabilität: Jede Phrase, bei der eine Bestimmung der Tonalität unmöglich ist, wird als *moduliert* gekennzeichnet. Als solche werden Phrasen bezeichnet, die entweder so falsch intoniert sind, dass ein tonales Zentrum nicht erkannt werden kann oder deren tonales Zentrum von dem der beiden vorangegangenen Takte abweicht (die Originalmelodie von Oura und Hatano bewegt sich eindeutig und ausschließlich in der Tonalität C-Dur).

Blockweise Zuordnung von Phrasen: Die nicht modulierten Teile der Wiedergaben werden nach den Gesangspausen in Phrasen eingeteilt. Für jede Phrase wird die beste Entsprechung in der Originalmelodie gesucht und dieser zugeordnet.

Identifizierung nach Halbtakten: Jede wiedergegebene Phrase und ihre Entsprechung der Originalmelodie werden jeweils in halben Takten nach Übereinstimmung in Kontur und/oder Rhythmus überprüft. Wenn sich eine Übereinstimmung in mindestens einem der beiden Parameter ergibt, wird dieser Halbtakt als *identifizierbar* gekennzeichnet. Bei Nicht-Übereinstimmung in beiden Parametern bekommt der Halbtakt der Wiedergabe das Label *nicht-identifizierbar*.

Bestimmung der Korrektheit: Die *identifizierbaren* Halbtakte werden danach hinsichtlich ihrer Übereinstimmung mit dem Original eingeteilt: Halbtakte, die exakt dem Original entsprechen sind *korrekte* Halbtakte. Ergeben sich jedoch Abweichungen in Rhythmus oder Tonhöhe, so wird der jeweilige Halbtakt als *modifiziert* kodiert. Zielinska und Miklaszewski (1992) bestimmen zudem noch die Art der Fehler in den modifizierten Halbtakten. Sie unterscheiden Fehler hinsichtlich der verwendeten Skala, der melodischen Kontur, der Intervalle und des Rhythmus.

Die *modulierten*, die *nicht-identifizierbaren*, die *identifizierbaren* und davon die *korrekten* und *modifizierten* (ggf. mit Differenzierung nach der Art der Modifikation) Halbtakte einer Wiedergabe werden aufsummiert und stellen die abhängigen Variablen des Versuches dar.

Diese abhängigen Variablen können dann in einer multivariaten Varianzanalyse mit Messwiederholung (MANOVA) auf Mittelwertsunterschiede in Abhängigkeit von den Faktoren Versuchsdurchgang, musikalische Expertise und Melodietyp (bei Oura & Hatano sowie Zielinska & Miklaszewski eine tonale ‚westliche‘ vs. eine modale japanische Melodie) untersucht werden.

Ergebnisse:

Da die Studie von Sloboda und Parker einen stark explorativen Charakter trägt, sind die von ihnen präsentierten Ergebnisse vor allem qualitativer Natur, während die beiden Folgestudien aus Japan bzw. Polen sich eher um eine Quantifizierung des Einflusses des Melodietypus und der musikalischen Expertise auf die Wiedergabeleistung bemühen.

Sloboda und Parker werteten jeweils sechs aufeinanderfolgende Wiedergaben der selben Melodie ihrer acht Versuchspersonen aus, insgesamt also 48 Wiedergaben. Zum Erstaunen der beiden Forscher fand sich unter diesen 48 Wiedergaben der simplen tonalen Melodie und trotz der musikalischen Vorbildung der Hälfte ihrer Versuchspersonen keine einzige

Wiedergabe, die vollkommen korrekt gewesen wäre. Trotzdem fanden sie eine hohe Korrelation zwischen den Wiedergaben und dem Original, die sich vor allem in der sehr guten Wiedergabe des Metrums (4/4), in der häufig erhaltenen Phrasenstruktur (AAB) und in der Wiedergabe einzelner Motive manifestierte. Zusätzlich war bei den musikalisch vorgebildeteren Versuchspersonen das harmonische Gerüst der Melodie in den Wiedergaben relativ gut zu erkennen.

Dennoch stellen Sloboda und Parker fest, dass mit zunehmender Zahl der Wiederholungen die Anzahl der wiedergegeben und der richtig wiedergegebenen Tönen zwar wuchs, jedoch das Verhältnis zwischen richtigen Tönen und allen wiedergegebenen Töne in etwa konstant blieb. D.h. die Wiedergaben wurden länger und ähnelten hinsichtlich Phrasenstruktur und Aufbau immer mehr dem Original, jedoch schien diese größere globale Ähnlichkeit auf Kosten von melodischen oder rhythmischen Details zu gehen. Sloboda und Parker zogen aus diesem Ergebnis den Schluß, dass das Memorieren von Melodien im Aufbau eines mentalen Modells vorrangig für die Melodiestructur besteht. Oberflächendetails würden nach dieser Annahme von den Versuchspersonen eher im Einklang mit anderen Parametern improvisiert, während die meiste kognitive Arbeit für das Erfassen und die Wiedergabe der Struktur verwendet würde. Damit rückt die Interpretation der Ergebnisse in die Nähe generativer Melodiegrammatiken, und die Autoren selbst weisen auf die Nähe dieser Vorstellung zum generativen Melodiesystem von Sundberg und Lindblom (1976) hin.

Oura und Hatano (1988) sowie Zielinska und Miklaszewski (1992) betrachteten zehn statt sechs konsekutive Wiedergaben ihrer 16 bzw. 31 Versuchspersonen. Durch diese Verlängerung des Experiments können zumindest Zielinska und Miklaszewski ein Anwachsen des Anteils der korrekten Halbtakte und eine zunehmende Fehlereliminierung bei den späteren Versuchsdurchgängen feststellen. Sie erklären deshalb, dass der Lernprozess über die Wiederholungen der Melodie in ihrem Experiment durch einen Wendepunkt gekennzeichnet sei, bis zu dem sowohl richtig als auch falsch wiedergegebene Töne zunehmen (vgl. Abb. 3.3-1). Nach diesem Wendepunkt nimmt die Rate der falschen Halbtakte ab, wobei gleichzeitig die Anzahl der insgesamt identifizierbaren Takte steigt. Dass dieser Wendepunkt in der Lernkurve im Experiment von Sloboda und Parker unentdeckt blieb, erklären die beiden polnischen Forscher mit der geringeren Anzahl (sechs) von Versuchsdurchgängen. Die Lage dieses Wendepunktes hängt nach der Interpretation von Zielinska und Miklaszewski von der Beschaffenheit der Melodie und den musikalischen Fähigkeiten bzw. Vorkenntnissen der Versuchspersonen auch hinsichtlich des Stils der Melodie ab. So wurde der Wendepunkt von den Absoluthörern, die Zielinska und Miklaszewski testeten,

früher erreicht und bei der modalen Melodie von beiden Versuchsgruppen später als bei der tonalen Melodie. Den positiven Einfluß von musikalischen Fähigkeiten konnten auch Ogawa et al. (1995, Exp. 2) in einem Wiedergabeexperiment mit sehr ähnlicher Methodik beobachten, in dem die musikalisch vorgebildeten Vpn. weniger Durchgänge bis zur korrekten Wiedergabe der Melodien benötigten und die Wiedergaben der einzelnen Durchgänge weniger Fehler enthielten als bei den musikalisch formal nicht vorgebildeten Versuchspersonen.

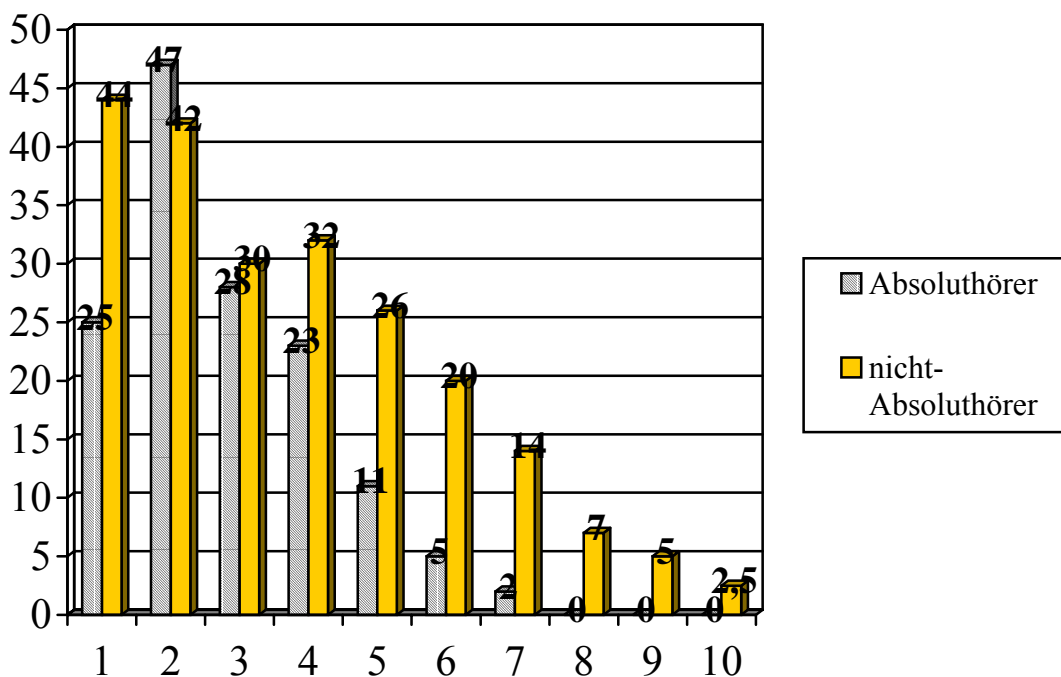


Abb. 3.3–1: Fehlerraten in Prozent (y-Achse) für Absoluthörer und nicht-Absoluthörer als Funktion der Anzahl der Wiedergaben (x-Achse) in Exp. 1 (tonale Melodie) von Zielinska und Miklaszewski (1992, S.103). (Eigene Darstellung)

Die Wendepunkte für die Entwicklung der Fehlerrate liegen jeweils beim zweiten Durchgang (47% und 42% Fehler) bei den Absoluthörern bzw. bei den Nicht-Absoluthörern. Die Absoluthörer erreichen 0% Fehler im achten Durchgang, die Nicht-Absoluthörern können nur bis 2,5% im zehnten Durchgang reduzieren. Die im Vergleich mit Oura & Hatano (1988) weit besseren absoluten Ergebnisse sind durch die Auswahl der Versuchspersonen – allesamt Musikstudenten – zu erklären.

Die Entdeckung des Wendepunktes in der Lernkurve bei Zielinska und Miklaszewski spricht jedoch nicht gegen den primären Erwerb der Melodiestruktur, den Sloboda und Parker gefunden hatten. Auch Zielinska und Miklaszewski konnten bei ihren Versuchspersonen

sonen – alle Musikstudenten – feststellen, dass die tonale Melodie in einzelne Phrasen segmentiert wurde und diese Phrasen nacheinander, zumeist beginnend mit der Anfangsphase des Stückes und mit Abweichungen in melodischen oder rhythmischen Details, wiedergegebenen wurden.

Auch die musikalisch vorgebildeten Versuchspersonen von Oura und Hatano zeigten eine ähnliche ‚musikalische‘ Verarbeitung und Segmentierung der Melodie. Jedoch konnten sie feststellen, dass der Lern- und Erinnerungsprozess bei musikalisch nicht vorgebildeten zumindest hinsichtlich der tonalen, ‚westlichen‘ Melodie teilweise deutlich anders verläuft: Nur etwa die Hälfte der musikalisch Unerfahrenen wurde überhaupt mit zunehmender Zahl der Wiederholungen auf die Gesamtheit der Takte bezogen besser. Im Gegensatz zur Segmentierung und Strukturierung von Phrasen bei den Musikern, zeigten viele Unerfahrene eine Fixierung auf die letzte Melodiephrase. Zusätzlich waren bei den Unerfahrenen häufig Modulationen in eine andere Tonart und Abweichungen in der Kontur zu erkennen. Die Tonhöhenfehler stellten bei den Unerfahrenen die größte Fehlergruppe dar, wobei Vereinfachungen von melodischen Phrasen, die sich bei den Musikern häufig gezeigt hatten, hier nur in sehr geringem Ausmaß zu beobachten waren. Ein hierzu komplementäres Ergebnis erhielten Ogawa und Kollegen in einem Wiedererkennensexperiment (Ogawa et al., 1995, Exp. 1). Nur die von ihnen getesteten musikalisch erfahrenen Versuchspersonen verwechselten signifikant häufig eine kurze japanische Melodie mit Melodievarianten, die dieselbe modale Struktur besaßen. Die Verwechslungen der Nicht-Musiker in diesem Experiment zeigten dagegen kein spezifisches Muster, welches auf die vorrangige Speicherung der Melodiestructur hingewiesen hätte.

Zusammengenommen deuten die Ergebnisse der Fehleranalyse von Oura und Hatano darauf hin, dass nur bei Versuchspersonen, die sich mindestens bis zu einem gewissen Grad aktiv mit Musik beschäftigt haben, eine Verarbeitung im Sinne von tieferliegenden musikalischen Strukturen bzw. von Generationsgrammatiken stattfindet. Dagegen zeigt die fehlende Segmentierung und Vereinfachung von melodischen Phrasen sowie das Modulieren in andere Tonarten bei Versuchspersonen, die nie in nennenswertem Maße aktiv musiziert haben, dass diese strukturellen Bestandteile von Melodien nicht als solche verarbeitet oder erinnert werden. Die Fehler, die diese Versuchspersonen im Gegensatz zu den getesteten Musikern zeigen, sind also eher ‚unmusikalisch‘ und dadurch wahrscheinlich schwerer zu systematisieren und zu prognostizieren.

Interessanterweise gilt dieser Vorteil der Musiker bei der Melodieverarbeitung jedoch nicht für die Wiedergaben der traditionellen japanischen Melodie, die weder eine deutliche Phra-

sensegmentierung enthielt noch markante rhythmische Eigenheiten und zusätzlich melodisch durchweg modal gehalten war. Hier zeigten sich weder eindeutige Unterschiede in den Arten der Fehler der beiden Versuchspersonengruppen noch eine signifikante Differenz in den absoluten Wiedergabeleistungen. Ein langsamer aber stetiger Fortschritt auf niedrigem Niveau war bei beiden Gruppen von Versuchspersonen zu verzeichnen. Offensichtlich sind also die Vorteile beim Verarbeiten und Erinnern der Melodiestructur auf Melodien in dem Stil begrenzt, in dem die Versuchspersonen aktive Erfahrungen gesammelt haben. Dies steht offensichtlich im Einklang mit den Erkenntnissen der Fehleranalysen von Cutietta & Booth (1996) und Kochmann (1923), bei denen die Versuchspersonen kurze und einfache volksliedähnliche, diatonische Tonfolgen teilweise mit ‚musikalisch nachvollziehbaren‘ Fehlern wiedergaben (s. Abschnitt 3.2.6.3).

Gestützt wird die Hypothese, dass Gedächtnisrepräsentationen von echten Melodien zunächst über eine Reduktion auf ein Melodieskelett funktioniert, von einer umfassenden Studie mit gänzlich verschiedener Methodik, die von Large, Palmer und Pollack (1995) durchgeführt wurde. Large et al analysierten Improvisationen von Pianisten über einfache Melodien und verglichen sie mit Reduktionen, die ein neuronales Netz der sogenannten Recursive-Auto-Associative-Memory-Architektur (RAAM) von denselben Melodien lieferte. Im Ergebnis zeigten sich große Ähnlichkeiten zwischen den Strukturen, die entlang der Improvisationen konstant blieben und den Reduktionen des trainierten Netzes. Large et al. schließen daraus, dass ein Reduktionsmechanismus mit großer Wahrscheinlichkeit an Gedächtnisprozessen zumindest mit unbekannten Melodien stark beteiligt ist. Ein Reduktionsmechanismus, wie der von Large und Kollegen vorgeschlagene könnte für die von Sloboda und Parker, Oura und Hatano sowie Ogawa et al. beobachtete präferierte Memorierung von Melodiestructuren bei musikalisch vorerfahrenen Personen zumindest in der Tendenz erklären.

3.3.2 Dimensionen der Speicherung von bekannten und unbekannten Melodien

Der Begriff ‚Dimensionen‘ wird im folgenden nicht in seiner mathematischen Definition gebraucht, sondern steht für die analytisch unterscheidbaren Aspekte von Melodien, wie Kontur, Rhythmus oder Tonalitätsbezug. Implizit wird dabei nicht die Unabhängigkeit der unterschiedlichen Dimensionen behauptet, wie dies beispielsweise für viele multivariate statistische Analyseverfahren vorausgesetzt wird. Den meisten Autoren der im folgenden

besprochenen Studien ist bewusst, dass die Konfiguration von Ereignissen oder Merkmalen in einer Dimension (z.B. Kontur) Auswirkungen auf die Wahrnehmung und Erinnerung der anderen Dimensionen (z.B. Rhythmus) haben kann. Dieses Wechselwirkungsverhältnis der verschiedenen Dimensionen einer Melodie wird von den neueren Akzentmodellen, die im anschließenden Abschnitt erörtert werden, genauestens untersucht und beschrieben.

Um festzustellen, welche Dimensionen bei der Speicherung von echten Melodien oder melodischen Passagen aus echten Musikstücken von Bedeutung sind, lassen sich grundsätzlich zwei verschiedene Ansätze ausmachen, nach denen die im folgenden besprochene Literatur eingeteilt ist:

Dem ersten Ansatz liegt die Überlegung zugrunde, dass die verschiedenen Dimensionen einer bekannten Melodie im Gedächtnis getrennt abgelegt sind bzw. wenige Dimensionen deutlich relevanter für die Identität oder das Erinnern einer bekannten Melodie – wie *Happy Birthday* – sind als andere. Das Erinnern einer bekannten Melodie wird in diesem Ansatz unter der Bedingung geprüft, dass einzelne Dimensionen oder Parameter selektiv manipuliert oder maskiert werden. Die Studien dieses Ansatzes der ‚manipulierten Dimensionen‘ werden im Abschnitt 3.3.2.1 behandelt.

Der zweite Ansatz verwendet statt bekannten Melodien für die Versuchspersonen neues melodisches Material als Versuchsstimuli, welches hinsichtlich mindestens einer Dimension variiert oder manipuliert wird. Neben der Bedeutung der verschiedenen musikalischen Faktoren werden in den meisten Experimenten dieses Ansatzes auch Parameter der Experimentalbedingungen (z.B. Retentionsintervall) oder subjektbezogene Faktoren untersucht. Abschnitt 3.3.2.2 fasst die Untersuchungen mit unbekannten Melodien zusammen.

Oberflächlich betrachtet lassen sich die beiden Ansätze als Untersuchungen zum Abruf von Melodien aus dem Langzeitgedächtnis (bekannte Melodien) und solche zum Speichern und Abruf aus dem Kurzzeitgedächtnis (unbekannte Melodien) klassifizieren. Wie jedoch die referierte Kritik am Mehrspeichermodell aus der Perspektive des sogenannten ‚Verarbeitungsebenen-Ansatzes‘ zeigt (z.B. Craik & Lockhardt, 1972), ist diese Einteilung in getrennte Gedächtnissysteme nicht unproblematisch. Zusätzlich ist für den Bereich des Melodiegedächtnisses im Besonderen noch nicht geklärt, wo die Grenze zwischen Kurzzeit- und Langzeitspeicher verläuft. Wie Berz (1995, S. 354) referiert, liegen die Angaben zum Kurzzeitgedächtnis für melodische bzw. musikalische Inhalte zwischen 2 und 180 Sekunden. Die Festlegung, ob eine melodische Gedächtnisaufgabe auf die Ressource des KZG oder des LZG zurückgreift, scheint damit eher von der Definition und dem theoretischen Hintergrund des jeweiligen Forschers abzuhängen, weshalb hier eine Einteilung an-

hand des Stimulusmaterials vorgenommen werden soll und nicht bezüglich der beteiligten Gedächtnissysteme.

3.3.2.1 Manipulation von Dimensionen bekannter Melodien

Der experimentelle Ansatz, bestimmte Parameter einer Melodie zu manipulieren oder zu maskieren, leitet sich aus Experimenten der Psychoakustik ab, die darauf abzielen, Grenzen zu ermitteln, innerhalb derer ein Signal noch als solches von den Versuchspersonen erkannt werden kann. Die typischen Fragestellungen dieser Experimente der 50er, 60er und 70er Jahre lauteten beispielsweise (vgl. Zwicker & Fastl, 1999): Wieviel weißes Rauschen verträgt ein Sprachsignal auf dem Signalweg, um noch verstanden zu werden? Durch welche anderen akustischen Ereignisse wird ein Ton in der Wahrnehmung maskiert? Wie wirken sich verschiedene Filter auf die Wahrnehmung von komplexen Klängen aus? Ziel dieser Untersuchungen war die Entwicklung von effizienten Übertragungs- und Wiedergabesystemen für die akustische Kommunikation (vgl. Terhardt, 1998).

Als einer der ersten übertrug Benjamin White (White, 1960) diesen Ansatz auf das Wiedererkennen bekannter Melodien, die auf verschiedene Weise manipuliert bzw. gestört wurden, wobei er sich an einer Pionierstudie von Heinz Werner (1926) orientierte. Whites zentrales Anliegen war es, diejenigen Transformationen oder Störungen zu identifizieren, die eine Melodie wiedererkennbar belassen. Dafür verwendete er zehn sehr bekannte amerikanische Volksmelodien, die jeweils zwölf Transformationen unterzogen wurden. Dazu gehörten die Veränderung der Intervallgrößen um verschiedene Faktoren, die Änderung der Tonreihenfolge oder Eliminierung des Rhythmus. Da White seine insgesamt nur neun Versuchspersonen immer wieder mit den selben Melodien unter jeweils verschiedenen Transformationen testete, sind Lerneffekte, die die Erkennensleistungen über die Versuchsdurchgänge hinweg beeinflussten, sehr wahrscheinlich. Dennoch kommt White zu Ergebnissen, die Folgestudien in den anschließenden 25 Jahren bestätigen konnten: Am wenigsten beeinträchtigen Transformationen das Wiedererkennen, die die relative Intervallgröße sowie die melodische Kontur intakt lassen. Sind diese Dimensionen intakt, so spielt der Rhythmus eine untergeordnete Rolle. Gleichzeitig konnten die Versuchspersonen die Melodien aufgrund des Rhythmus alleine nur in 33% der Fälle identifizieren. Interessanterweise hatte die Länge des Melodieausschnitts (6 vs. 24 Töne) keinen besonderen Einfluß auf das Erkennen der Melodie, was bedeutet, dass bei bekannten Melodien schon

sehr wenige Töne zum Erkennen ausreichen, wenn bei diesen Tönen die Kontur und die (relativen) Intervallinformationen intakt sind. Zusätzlich musste White z.T. große Unterschiede zwischen den Leistungen seiner Versuchspersonen feststellen. Schließlich vermutete er zusätzlich eine Wechselwirkung zwischen der Art der Transformation und der Struktur der Melodie. Gesetze für die Wiedererkennbarkeit von Melodien scheinen also neben den spezifischen Transformationen in nicht geringem Maße auch von Subjektfaktoren und den Eigenheiten des Stimulusmaterials abhängig zu sein.

Die Bedeutung von Whites früher Studie lässt sich an der Zahl der Zitierungen erkennen, die seine Veröffentlichung in den Fortführungen dieses Ansatzes mit eher psychologischer oder musikanalytischer Zielrichtung erfuhr.

Dowling und Fujitani vergleichen im Experiment 2 ihrer bekannten Studie von 1971 die Bedeutung von Kontur- und Intervallinformationen beim Wiedererkennen von bekannten Melodien mit dem Einfluss dieser Faktoren bei kurzen Tonfolgen (Dowling & Fujitani, 1971, Exp. 1; s. Kap. 3.2), indem sie eine an White angelehnte Prozedur verwenden. Tatsächlich sind nach ihren Ergebnissen im Gegensatz zu kurzen Tonfolgen bei echten Melodien nicht nur die Konturinformationen von Bedeutung, sondern auch die relative Intervallgröße und die exakten Intervalle haben einen signifikanten Einfluss auf die Wiedererkennungseistung. Dowling und Fujitani interpretieren dies im Sinne unterschiedlicher Parameter, die in KZG und LZG gespeichert werden: Im LZG wäre demnach die exakte Intervallgröße abgelegt, während sich das KZG nur mit der Kontur einer Melodie zum Wiedererkennen behelfen könnte.

Diana Deutsch (1972) überprüfte innerhalb dieses Ansatzes zusätzlich die Dimension des Tonchromas, also die Zugehörigkeit zu einer über mehrere Oktaven hinweg generalisierbaren Tonklasse. Sie verteilte in ihren Experimentalstimuli die Töne einer bekannten Melodie über mehrere Oktaven, wobei das Tonchroma jeweils erhalten blieb. Die Versuchspersonen, die diese oktavversetzte Version zu hören bekamen, zeigten im Ergebnis keine bessere Wiedererkennungseistung als eine Kontrollgruppe, die nur den Rhythmus hörten. Nach Deutschs Interpretation kann somit der Verarbeitungs kanal, in dem Melodien verarbeitet werden, keine Oktaven generalisieren, d.h. keine Information aus dem Tonchroma entnehmen. Die Verarbeitung und Speicherung von Melodien funktioniere deshalb vermutlich im Einklang mit der Verarbeitungstheorie von Deutsch (1969) als sukzessive Verarbeitung von Intervallen. Darüber hinaus stellte Deutsch fest, dass verbale Zusatzinformationen und Vorwissen beim Wiedererkennen der Melodie sehr hilfreich waren. Dieses Ergebnis

konnte Dowling ein Jahr später (1973b) bestätigen. In einer Experimentalreihe zur Stimmtrennung konnten die Versuchspersonen mit der Vorinformation, um welche Melodie es sich handele, diese sehr viel schneller aus einem Melodiengemisch identifizieren als ohne diese Information (Exp. 3). Offenbar ermöglicht diese Form von gelenkter Aufmerksamkeit eine andere kognitive Strategie in Form einer ‚aktiven Suche‘ in den dargebotenen Stimuli.

In der Folge konnten jedoch zwei Studien Deutschs Ergebnis, dass das Tonchroma für das Wiedererkennen von bekannten Melodien keine Rolle spiele, korrigieren bzw. differenzieren. Dowling und Hollombe (1977) sowie Idson und Massaro (1978) konnten in ihrer direkten Bezugnahme auf Deutschs Experiment zeigen, dass das Tonchroma sehr wohl zum Wiedererkennen von bekannten Melodien mit oktavversetzten Tönen benutzt wird, jedoch nur unter der Voraussetzung, dass die Kontur – also das Auf und Ab der Melodie – bei den Oktavversetzungen erhalten bleibt. Dennoch werden selbstverständlich auch unter diesen Bedingungen – identische Kontur und identische Tonchromata – die Volksmelodien signifikant schlechter wiedererkannt als in ihrer Originalgestalt.

Im Gegensatz zu Deutsch wurden in diesen beiden Studien eine ganze Reihe von bekannten Volksmelodien als Testmaterial verwendet, so dass die Bedeutung der Kontur sowie des Tonchromas in Verbindung mit der Kontur für einen größeren Stimulusbereich generalisiert werden konnte. Jedoch stellten beide Studien übereinstimmend eine bedeutsame Interaktion zwischen den einzelnen Melodien und der Art der ‚Maskierung‘ der Melodien fest. So wurden beispielsweise die ersten vier Phrasen des Liedes „Three blind mice“ in 94% der Durchgänge mit erhaltener Kontur und erhaltenem Chroma identifiziert, während unter der gleichen Bedingung in keinem einzigen Versuchsdurchgang (0%) „When the saints go marching in“ erkannt wurde (Dowling & Hollombe, 1977, S. 63). Genau wie White (1960) kommen deshalb sowohl Dowling und Hollombe als auch Idson und Massaro zu dem Schluss, dass neben der Art und dem Grad der Manipulation oder Störung für die vollständige Beschreibung des Prozesses des Wiedererkennens auch Eigenschaften der Struktur der Melodien sowie Subjektfaktoren von großer Bedeutung sind. Über die einflussnehmenden strukturellen Eigenschaften stellen die Autoren hier jedoch keine Vermutungen an. Als Subjektgröße beeinflusse auf jeden Fall die musikalische Expertise den Vorgang des Wiedererkennens, schreiben Dowling und Hollombe (S. 63): Je geringer die musikalische Vorbildung, desto größer sei die Bedeutung der melodischen Kontur für das Wiedererkennen.

Schon White (1960) entdeckte, dass das Wiedererkennen bekannter Melodien allein aufgrund von rhythmischer Information seltener erfolgreich ist, als wenn die gleiche Aufgabe mit intakter melodischer Information gestellt wird. Die daraus resultierende Annahme, dass die Tonhöhenfolge (Melodie) eines bekannten Liedes mehr Information für das Wiedererkennen trägt und damit ein stärkerer *Gedächtniscue* ist als der Rhythmus des Liedes, konnten 37 Jahre später Hébert und Peretz (1997) bestätigen. Auch die unwillkürliche selektive Aufmerksamkeit der getesteten Versuchspersonen schien sich in den Wiedererkennungsaufgaben eher dem melodischen als dem zeitlich-rhythmischen Muster zuzuwenden (Exp. 2). Dennoch wurden Melodien mit intaktem Tonhöhenverlauf und Rhythmus noch deutlich besser erkannt als die reinen melodischen Varianten (Exp. 1); d.h. auch der Rhythmus – wenn auch in geringerem Maße als die Tonhöhenfolge – trägt zum Wiedererkennen bekannter Melodien bei. Diesen geringeren Einfluß des Rhythmus halten die beiden kanadischen Forscher zumindest teilweise für kulturell begründet: Die Volkslieder seien besser durch die Tonhöhenfolge als durch den Rhythmus unterscheidbar. Außerdem seien die Versuchspersonen von der stärker melodisch als rhythmisch orientierten westlichen Musikkultur geprägt.

Auch die bereits von Deutsch (1972) sowie von Dowling und Hollombe (1977) beschriebene Effektivität von verbalen Zusatzinformationen konnten Hébert und Peretz bestätigen. Durch diese Zusatzinformationen werde ein *Top-Down-Prozess* möglich, in dem eine mentale Repräsentation mit dem aktuellen sensorischen Input verglichen werde. Dabei sei es besonders für das Wiedererkennen des Rhythmus von Bedeutung, dass dieser Vergleich vor oder zu Beginn der Präsentation des sensorischen Inputs erfolge, da dieser nicht unbedingt erfolgreich für einen Vergleich im KZG enkodiert werden könne. Anders formuliert bedeutet dies, dass vor allem verbale **VOR**informationen für das Wiedererkennen von bekannten Melodien effektiv sind, bei denen mindestens eine Dimension manipuliert oder durch Störreize maskiert ist.

3.3.2.2 Speicherung von unbekanntem melodischen Material

Die in diesem Abschnitt referierten Studien sind in Hinblick auf das experimentelle Paradigma, die verwendeten Stimuli und den internen Bezugnahmen weniger homogen als die in den vorangegangenen Abschnitten diskutierten Arbeiten.

Gemeinsam ist den im folgenden dargestellten Untersuchungen die Verwendung von unbekanntem musikalischen Material, welches melodische Qualität besitzt und einen Aus-

schnitt aus einem echten Musikstück darstellt, womit ein deutlich höherer Grad an ökologischer Validität erreicht wird als bei analogen Studien mit kurzen Tonfolgen. Trotz der Unterschiedlichkeit der methodischen Herangehensweisen liefern die Ergebnisse der besprochenen Studien in der Zusammenschau gute Hinweise auf die Funktionsweise des melodischen oder musikalischen Gedächtnisses unter wirklichkeitsnahen Bedingungen.

Die Experimente 5 bis 7 der Studie von Dowling, Kwak und Adreus (Dowling et al., 1995) stehen eindeutig in dem für Jay Dowling typischen und an kurzen Tonfolgen erprobten Forschungsansatz (Wiedererkennensparadigma, Vergleiche von Melodien/Tonfolgen auf unterschiedlichen Transpositions-niveaus, Bedeutung von Kontur vs. Intervallen etc.), wie er im Kapitel 3.2 referiert wurde. Dennoch gelangen Dowling und Mitarbeiter durch die Verwendung von Ausschnitten aus echten Melodien, die für die Versuchspersonen z.T. bekannt bzw. unbekannt sind, zu einer Reihe von neuen Erkenntnissen: Die Versuchspersonen können sehr viel besser exakte Transpositionen einer Melodie und Transpositionen mit leicht verschiedener Kontur voneinander unterscheiden als dies mit künstlichen Tonfolgen der Fall war. Die Autoren sehen diese deutlich besseren Unterscheidungsleistungen in den u.a. durch den Rhythmus sehr viel reicheren *cues* der echten Melodien begründet.

Im Vergleich mit den bekannten echten Melodien sind die Gedächtnis- und Unterscheidungsleistungen aller Versuchspersonen mit unbekannten Melodien – wie zu erwarten – signifikant schlechter. Dennoch stellen Dowling und Kollegen fest, dass die musikalische Vorbildung offenbar die Verarbeitung von neuen Melodien beeinflusst. Denn vor allem ihre musikalisch wenig erfahrenen Versuchspersonen verlassen sich stärker auf die Informationen der absoluten Tonhöhe und der Tonart, während die Versuchspersonen mit größerer Expertise im Musikbereich besser mit Transpositionen der ihnen unbekannten Melodien umgehen können und diese deutlich besser erinnern. Dies lässt sich im Licht der Ergebnisse von Sloboda & Parker (1985) bzw. der entsprechenden Nachfolgestudien (s.o., Abschnitt 3.3.1) vielleicht als Hinweis auf die Wahrnehmung und Speicherung der reduzierten strukturellen Informationen einer Melodie lesen, die ja vor allem bei musikalisch vorgebildeten Personen ausgeprägt sind.

In einer 2002 publizierten Experimentalreihe gehen Dowling und Kollegen (Dowling et al., 2002) noch einen Schritt weiter, indem sie jeweils den Anfang echter und einfacher Kla-

vierstücke (Menuette, Walzer, Ländler von Mozart, Beethoven, Schubert) als Stimuli verwenden. Die jeweils erste oder zweite Phrase zu Beginn dieser Klavierstücke sollten die Vpn. mit einer Folgephrase, die zwischen 4 und 30 Sekunden später erklingt (Retentionsintervall), vergleichen und angeben, ob es sich um eine exakte Wiederholung, eine ähnliche Wiederholung oder eine vollkommen verschiedene Phrase handelte.

Die Forscher fanden einige interessante, wenn auch nicht ganz unerwartete Haupteffekte der kontrollierten Einflussfaktoren.

Für die erste Phrase gleich zu Beginn des Stückes zeigten die Versuchspersonen signifikant bessere Gedächtnisleistungen als für die zweite Phrase. Die Eröffnungsphrase eines Stückes scheint also eine gewisse Wahrnehmungspriorität zu haben.

War zwischen der Referenzphrase und der Folgephrase keine Musik zu hören, sondern Stille eingefügt, so steigerte dies die Gedächtnisleistung der Versuchspersonen. Daraus lässt sich schließen, dass auch die zuletzt gehörte musikalische Phrase gut im Gedächtnis bleibt.

Die Versuchspersonen mit musikalischer Vorerfahrung schnitten nur bei einem der fünf Experimente besser ab als ihre studentischen Kollegen ohne einschlägige Vorbildung.

Für interessanter noch als diese Haupteffekte befanden die Autoren die Interaktionseffekte, die Auskunft über das komplexe Zusammenspiel und über die zeitliche Entwicklung der Einflussfaktoren geben. So wurde die Unterscheidung zwischen identischen und ähnlichen Wiederholungen mit zunehmendem zeitlichen Abstand der Phrasen deutlich besser, jedoch nur unter der Bedingung, dass in der Zwischenzeit der normale Fortgang des Stückes zu hören war. Bekamen die Versuchspersonen zwischen den beiden zu vergleichenden Phrasen Stille oder nur ein rhythmisches Muster zu hören, steigerte sich die Unterscheidungsleistung nicht mit zunehmender Zeit. Dass die Unterscheidung von identischen und ähnlichen Phrasen mit der Zeit besser wird, lässt auf eine fortschreitende Präzisierung des Gedächtnisinhalts schließen, welche im Gegensatz zu vielen empirisch ermittelten Gedächtnisprozessen – z.B. für verbale Inhalte – steht, die eher durch eine ausgeprägte Vergessenskurve beschrieben werden. Dowling und Mitarbeiter schließen deshalb, dass beim Hören von echter Musik gleichzeitig eine Verarbeitung des gerade Gehörten stattfindet. Diese Verarbeitung sei ein automatischer, nicht von Aufmerksamkeit gesteuerter Prozeß, den die Autoren mit der Modellvorstellung des impliziten Gedächtnisses in Verbindung bringen (z.B. Schacter, 1987). Dennoch hatte auch die willkürliche Fokussierung der Aufmerksamkeit einen erheblichen Anteil an den Gedächtnisleistungen in den insgesamt fünf Experimenten dieser Reihe: Wurden die Versuchspersonen zu Beginn nicht instruiert, dass es sich

um einen Gedächtnistest handle – weswegen sie folglich den beiden ersten Phrasen keine besondere Aufmerksamkeit schenkten – so war keine Änderung der Unterscheidungsleistungen mit der Zeit zu beobachten, sodass vermutlich kein besonderer Verarbeitungsprozess für die ersten beiden Phrasen zum Tragen kam.

Insgesamt halten Dowling und Mitarbeiter die Leistung in Gedächtnistests mit echter Musik durch eine Mischung aus impliziten automatischen und expliziten willkürlichen Prozessen begründet. Die Gedächtnisrepräsentationen musikalischer Bestandteile unterliege dabei mit zunehmender Zeit einem kontinuierlichen und qualitativen Wandel und werde dabei immer präziser. Allerdings sei noch ungeklärt, welche Rolle das musikalische Material spiele, welches im Retentionsintervall erklingt. Fest stehe aber, dass dieses Material auch bei musikalisch nicht vorgebildeten Versuchspersonen eine große Bedeutung für die Gedächtnisprozesse habe.

Insgesamt demonstrieren Dowling und Kollegen mit dieser Studie, wie sie durch die Verwendung von echtem melodischen Material zu qualitativ neuen Ergebnissen gelangen, die bei der Untersuchung des Gedächtnisses für kurze Tonfolgen nicht zu beobachten waren. Dazu zählen die Unterscheidung von automatischen und willkürlichen Gedächtnisprozessen und der kontinuierliche sowie qualitative Wandel der Gedächtnisrepräsentation einer bestimmten melodischen Phrase im fortlaufenden Prozess des Hörens.

Einen methodisch sehr ähnlichen Ansatz wählten William Kauffman und James Carlsen in ihrer Studie zum Wiedererkennen von Ausschnitten aus unbekannten klassischen Kompositionen (Kauffman & Carlsen, 1989). Wie bei Dowling et al. (2002) stand hier die Absicht im Vordergrund, Ergebnisse für einen Gedächtnistest zu erhalten, der möglichst nahe an echten Musikrezeptionssituationen angelehnt ist und so als ökologisch valide gelten kann. Die Versuchspersonen bekamen den Anfang eines Musikstückes der klassischen oder romantischen Epoche zu hören, dessen Anfangsperiode (zwei Phrasen mit deutlicher Schlusskadenz) das zu behaltende Zielobjekt darstellte. Die Versuchspersonen hörten im Anschluß an dieses Zielobjekt je nach Versuchsbedingung für 0 bis 180 Sekunden (Retentionsintervall) den normalen Fortgang des Stückes. Danach wurde den Versuchspersonen ein einzelner Ausschnitt der Komposition präsentiert, der entweder identisch mit dem Zielobjekt war oder einen vollkommen verschiedenen, aber in etwa gleich langen Teil der selben Komposition darstellte. Die Aufgabe der Versuchspersonen bestand darin anzugeben, ob sie diesen Ausschnitt zuvor schon gehört hatten.

Da die Versuchspersonen von Kauffman und Carlsen im Gegensatz zur späteren Studie von Dowling et al. (2002) nur identische und ganz verschiedene, jedoch keine ähnlichen oder variierten melodischen Passagen zu unterscheiden hatten, sind die Ergebnisse der beiden Studien nur bedingt miteinander vergleichbar. Insgesamt scheint die Aufgabe von Kauffman und Carlsen weit einfacher gewesen zu sein: 84% aller Antworten waren korrekt, wobei die Personen mit größerer musikalischer Vorbildung signifikant bessere Leistungen zeigten. Obwohl Kauffman und Carlsen dies nicht berichten, mag bei den Ergebnissen dieser Personengruppe ein *ceiling*-Effekt aufgetreten sein. Im Gegensatz zu Dowling et al. (2002) stellten Kauffman und Carlsen keinerlei Verbesserungen der Unterscheidungs- bzw. Gedächtnisleistungen mit zunehmendem zeitlichen Abstand von Ziel- und Vergleichsausschnitt (Retentionsintervall) fest. Im Gegenteil: Die Daten passen sich sehr gut der negativen exponentiellen Vergessenskurve für das KZG an, die beispielsweise bei Anderson (1995) sehr gut dargestellt ist. Die Vergessenskurven für Musikexperten und Personen mit geringerer musikalischer Vorbildung erwiesen sich als sehr ähnlich, auch wenn sie absolut gesehen auf einem unterschiedlichen Leistungsniveau lagen. Deshalb vermuten die beiden Forscher eine Überlegenheit der Musikexperten beim Enkodieren der melodischen Periode, statt einer simplen größeren Kapazität des (musikalischen) KZGs.

Das Auftauchen der in der Gedächtnisliteratur (z.B. Anderson, 1995) wohl bekannten Vergessenskurve lässt die Autoren schließen, dass die Aufgabe nur das KZG der Versuchspersonen bemühte. Obwohl das längste Retentionsintervall von drei Minuten eigentlich weit außerhalb der üblichen Grenzen des KZGs liegt (Kluwe, 1990, S. 129-130), sprechen die Autoren von einem erweiterten KZG, welches durch effektives Chunking und das Zuweisen von Bedeutung zu der präsentierten echten Musik möglich werde (Kauffman & Carlsen, 1989, S. 11). Insofern sei es ein erheblicher Unterschied, ob konstruiertes, weitgehend ‚musikalisch sinnfreies‘ Material oder existierende Musik als experimentelle Stimuli verwendet würden.

Aufgrund der realitätsnahen experimentellen Methode, sind die Befunde zu Vergessenskurve und musikalischer Expertise mit Sicherheit für die Beschreibung des Rahmens, in dem sich Prozesse des musikalischen Gedächtnisses abspielen, von Bedeutung. Wie gut dieses ‚erweiterte KZG‘ für echte Musik arbeitet, zeigte auch eine Nachfolgestudie in ähnlichem Design von Kauffman (Kauffman, 1990), in der Kinder im Alter von 7-8 bzw. 10-11 Jahren als Versuchspersonen untersucht wurden. Im Gegensatz zu den Erwachsenen im Experiment von 1989 war das Retentionsintervall in diesem Experiment nicht mit Musik gefüllt, sondern eine einfache Pause von 0, 10, 20 oder 40 Sekunden. Zu Kauffmans Über-

raschung zeigten die Kinder keine schlechteren Unterscheidungsleistungen nach dem langen Retentionsintervall als im sofortigen Anschluss der Zielphrase. Offensichtlich war die Aufgabe für die Kinder zu leicht, sodass sich ein *ceiling*-Effekt in den Daten zeigte. Ein signifikanter Vorteil war bei den älteren Kindern dennoch festzustellen. Festhalten lässt sich, dass in dem Fall, in dem keine störende Musik im Retentionsintervall erklingt, schon Kinder von 7 und 8 Jahren eine stabile Gedächtnisrepräsentation von Ausschnitten aus echter Musik bis mindestens 40 Sekunden erreichen können, die zur Bewältigung einer Unterscheidungsaufgabe ausreicht.

Leider berücksichtigen Kauffman und Carlsen in beiden Versuchen keine Parameter der musikalischen Struktur, die Aufschluss darüber geben könnten, welche strukturellen Eigenschaften einer melodischen Passage für deren Erinnerbarkeit verantwortlich sind. Durch die Einbeziehung von Vergleichsausschnitten, die in bestimmten Dimensionen manipuliert sind, hätte dies miterhoben werden können.

Einen Zusammenhang zwischen Parametern der musikalischen Struktur und der Erinnerbarkeit von *Jazzinstrumentals* (von Bebop bis Free Jazz) konnte Russell (1987) herstellen. Allerdings verzichtete er weitgehend auf die schwierige und nicht eindeutige Bestimmung der strukturellen Eigenschaften der von ihm gewählten Ausschnitte und ließ seine Versuchspersonen die getesteten Stücke auf insgesamt sieben Dimensionen einschätzen, darunter Komplexität, Tempo, Ensemblegröße, Melodiösität und Erinnerbarkeit. Diese Methode der Bestimmung struktureller Variablen durch subjektive Einschätzungen der Versuchspersonen entstammt den Fortentwicklungen des Ansatzes der sogenannten *New Experimental Aesthetics* (vg. Berlyne, 1971), wie sie in Deutschland beispielsweise von Rainer Niketta eingesetzt wurde (z.B. Niketta, 1990).

Bei der Korrelation der subjektiven Bewertungen untereinander konnte Russell einen engen Zusammenhang zwischen den geschätzten Dimensionen Bekanntheit mit musikalischem Stil des Ausschnitts, Melodiösität und ästhetischem Gefallen ausmachen. Die Stücke, die auf diesen Dimensionen hoch bewertet wurden, wurden gleichzeitig als weniger komplex beurteilt und insgesamt von den Versuchspersonen am besten erinnert. Interessanterweise ergab sich keinerlei Zusammenhang zwischen der von den Versuchspersonen eingeschätzten Erinnerbarkeit der Ausschnitte und der tatsächlichen Wiedererkennungsrate. Russell schlussfolgert deshalb, dass seine Versuchspersonen der subjektiven Dimension Erinnerbarkeit („memorability“) keine sinnvolle Bedeutung zuweisen konnten.

Dagegen korrelierten die subjektiven Maße von Ensemblegröße und Tempo mit den von Russell objektiv (Anzahl Instrumente bzw. Schläge pro Minute) bestimmten. Es schien den Versuchspersonen also leicht möglich, diese objektiven Parameter nachzuvollziehen. Einleuchtend ist auch die Korrelation zwischen subjektiver Komplexität und objektivem Tempo, die Russell ermittelte.

Im Gegensatz zu den meisten zuvor referierten Studien zeigten die Versuchspersonen mit größerer musikalischer Vorbildung keine besseren Erinnerungsleistungen. Da gleichzeitig die Bekanntheit mit dem Stil des Jazzstückes hoch mit der Gedächtnisleistung korrelierte und nur zwei der über 400 Vpn. Jazz als am meisten präferierte Musik angaben, ist zu vermuten, dass die musikalische Kompetenz der Versuchspersonen sich vor allem auf Popmusik oder klassische Musik bezog und nicht zu einer besseren Verarbeitung und Speicherung der Jazzstücke beitragen konnte. Dies würde sich mit den Ergebnissen von Oura und Hatano (1988; s. 3.3.1) decken, die auch einen Zusammenhang aufzeigen konnten zwischen den Erfahrungen, die Versuchspersonen im aktiven Umgang mit einem musikalischen Stil gesammelt haben, und den Gedächtnisleistungen für Musik diesen Stiles. Diese Vermutung wird zusätzlich durch den korrelativen Zusammenhang zwischen Gefallen, Melodiösität, Wiedererkennensrate und die gleichzeitige negative Korrelation mit der subjektiven Komplexität gestützt. Vielleicht wurden gerade die Stücke am besten erinnert, die eine höhere Ähnlichkeit mit der von den meisten Versuchspersonen präferierten Popmusik zeigten (Erhebung der Präferenzen über einen Fragebogen) und dadurch im Durchschnitt auch mehr gefielen.

Neben der Vertrautheit mit dem Musikstil der musikalischen Passage, die behalten werden soll, ist auch die Möglichkeit zur Rekodierung bzw. die Verbindung der gehörten Musik mit *cues* aus anderen Wahrnehmungsmodalitäten ein Faktor, der die Erinnerung an Ausschnitte aus echter Musik deutlich verbessern kann. In einer 1978 veröffentlichten Studie konnten Delis et al. zeigen, dass einminütige Ausschnitte aus symphonischen Werken dann besser behalten werden, wenn die Versuchspersonen einen konkreten Titel (d.h. eine programmatische Überschrift, wie z.B. „Winter Forrest“ oder „Ocean Voyage“) zu den akustischen Exzerpten genannt bekamen, mit dessen Hilfe sie eine mentale bildliche Vorstellungen zur gehörten Musik generieren konnten. Abstrakte Titel (z.B. „Refuge in Truth“, „Philosophical Question“) regten dagegen weit weniger lebhaft Vorstellungen bei den Versuchspersonen an und zogen so schlechtere Erinnerungsleistungen nach sich.

Delis und Kollegen legen mit diesen Ergebnissen nahe, dass das Verbinden von Musik mit bildhaften Vorstellungen bei der Enkodierung zu einer feststellbaren höheren Gedächtnisleistung für die zu behaltende Musik führt. Dies stellen die Autoren in bewusste Analogie zu Erkenntnissen zum Behalten verbaler Inhalte, welches durch visuelle Vorstellung (z.B. Paivio, 1969) oder Einbettung in einen sinnvollen Kontext (z.B. Bransford & Johnson, 1972) verbessert werden kann. Es lassen sich aus der jüngeren Vergangenheit viele ähnliche experimentelle Ergebnisse finden, die verbesserte Gedächtnisleistungen durch die Verbindung der Experimentalstimuli mit anderen, auch selbst generierten *cues* belegen (z.B. Slamecka & Graf, 1978; Slamecka & Fevreiski 1983; Eich & Metcalfe 1989). Dazu gehören natürlich auch Gedächtnismethoden, wie die bekannte „Method-of-loci“ (vgl. Anderson, 1995, S. 281). Im Lichte dieser Erkenntnisse aus der allgemeinen Gedächtnisforschung ist das Ergebnis von Delis und Kollegen also nur folgerichtig für den Bereich der Musik.

Zusätzlich zum signifikanten Einfluss der mit dem Musikausschnitt verbundenen Titel verzeichneten Delis et al. aber auch signifikante Unterschiede in der Behaltensleistung, die auf die gewählten Musikausschnitte selbst zurückzuführen waren. Die Autoren vermuten die verschiedenen in den verwendeten Ausschnitten zu hörenden „easy-to-comprehend-themes“ (Delis et al., 1978, S. 217) als Ursache für diese Unterschiede. Auch Delis und Ko-Autoren verweisen damit auf strukturelle Eigenschaften des melodischen Materials, welches das Wiedererkennen begünstigen oder erschweren kann. Leider gehen die Autoren der Studie jedoch nicht weiter auf diese strukturellen Komponenten der Erinnerungsleistung ein.

3.3.3 Bedeutung von Akzenten und melodischer Erwartung für das Melodiegedächtnis

Das Besondere an den in diesem Abschnitt referierten Studien ist die detaillierte Betrachtung des Aufbaus der melodischen oder musikalischen Struktur der zu erinnernden Stimuli. Während in den vorangegangenen Abschnitten dieses Kapitels vor allem Faktoren der experimentellen Anordnung (z.B. Anzahl der Präsentationen, Länge des Retentionsintervalls), subjektbezogene Faktoren (z.B. musikalische Expertise, Bekanntheit mit Stil), Faktoren der globalen Bedeutung oder einzelner musikalischer Dimensionen (z.B. Tonchroma, Intervallinformationen) im Vordergrund standen, suchen die nachfolgenden Studien auf einer tieferen Analyseebene nach speziellen musikalischen Faktoren. Es sind hier die Muster der musikalischen Ereignisse in den einzelnen Dimensionen in ihrer Wirkung auf

das Melodiegedächtnis von Interesse. Die dahinterliegende Forschungsfrage lautet: Wie wirkt sich der Aufbau einer Melodie (oder einer melodischen Passage) auf das Behalten dieser Melodie aus? Um diese Frage beantworten zu können, sind Schwierigkeiten auf mehreren Stufen zu meistern:

Es muss ein sinnvolles und angemessenes Repräsentationssystem für die experimentellen Stimuli, i.d.R. Melodien, entworfen werden, das die relevanten Unterschiede und Identitäten sowie die spezifischen Eigenschaften (Konfigurationen) von verschiedenen Melodien adäquat beschreibt. Im Jargon des *Music-Information-Retrieval*-Paradigmas formuliert: Es müssen die besten Deskriptoren zur Beschreibung von Melodien bzw. des Melodiegedächtnisses gefunden werden.

Es muss eine Theorie oder zumindest eine Heuristik existieren, die aufgrund von vorliegenden Konfigurationen einer bestimmten Melodie in diesem Repräsentationssystem Vorhersagen über die globale Erinnerbarkeit der Melodie oder einzelner Bestandteile dieser Melodie im Experiment erlaubt.

Es muss ein experimentelles Verfahren/Paradigma existieren, mit dem die Erinnerungsleistung der Versuchspersonen an diese Melodie empirisch ermittelbar ist.

Es muss eine Metrik entworfen werden, die die Erinnerungsleistungen in Relation zur Originalmelodie quantifiziert, sodass die Voraussagen der Theorie (mit statistischen Verfahren) überprüfbar werden.

Die Ansätze der verschiedenen Autoren, die im folgenden referiert werden, unterscheiden sich voneinander auf einer oder mehreren der genannten vier Stufen. Da die Lösungen und Modelle auf den einzelnen Stufen innerhalb eines Ansatzes meist voneinander abhängen bzw. aufeinander bezogen sind, erscheint es sinnvoll, die zu einem Ansatz gehörenden Studien zusammen zu referieren und danach die Qualität der Ergebnisse der verschiedenen Ansätze in einer Zusammenschau zu vergleichen. Die meisten Ansätze sind geprägt von den Veröffentlichungen von einem oder wenigen Autoren, die folglich auch bei der hier gewählten Benennung der Ansätze berücksichtigt werden.

3.3.3.1 Das melodisch-rhythmische Konsonanzmodell von Caroline B. Monahan

Die Idee hinter dem Modell der melodischen und rhythmischen Konsonanz ist die, dass innerhalb der verschiedenen musikalischen Dimensionen Akzente auf verschiedene Weise

entstehen. Fallen diese Akzente zeitlich zusammen, so bekommt das musikalische Ereignis zu diesem Zeitpunkt (d.h. der Ton) ein besonderes Gewicht für die menschliche Wahrnehmung und wird dadurch besser enkodiert und länger gespeichert.

Monahan (Monahan & Carterette, 1985, S. 3) nennt vier Quellen für Akzente in erklingender (einstimmiger) Musik: Die Lautstärke einer Note (lautere Töne haben einen Akzent), die Notendauer (längere Töne haben Akzent), die Beziehung zur Tonalität (Grundton und Quinte zu erklingendem Akkord haben Akzent) und Änderung in der Tonhöhenfolge (Konturumkehrungen und Sprünge haben Akzent). Empirisch untersucht Monahan jedoch nur den Einfluß bzw. die Wechselwirkung von Tondauern und Änderungen in der Tonhöhenfolge.

Rhythmuskonzept:

Die zeitlichen bzw. rhythmischen Dimensionen besitzen bei der Verarbeitung und Speicherung von kurzen Tonfolgen Priorität, wie Monahan und Carterette (1985) in einer Studie zeigen konnten, die das Verfahren der Multidimensionalen Skalierung zur Ermittlung des Wahrnehmungsraumes von kurzen Tonfolgen einsetzte. Danach wurde die Ähnlichkeitsbewertung der kurzen tonalen Folgen vor allem durch das Metrum und das Setzen der rhythmischen Akzente bestimmt. Monahan spricht deshalb von „salience of temporal accenting and grouping over pitch-level or tonal grouping“ (Monahan, 1993, S. 151). Ihr Rhythmuskonzept spielt deshalb eine besondere Rolle im Konsonanzmodell und ist von ihr auch entsprechend differenziert entworfen worden. Die Grundzüge dieses Rhythmuskonzeptes seien hier kurz zusammengefasst und für eine spätere eigene Anwendung aufbereitet.

Für die Repräsentation von Tonfolgen und Melodien bzw. für die Ermittlung der Akzente von Melodien verwendet sie für den Bereich der Tondauern vor allem empirische Erkenntnisse von Povel (z.B. Povel, 1981; Povel & Okkerman, 1981; Povel & Essens, 1985) hinsichtlich der wahrgenommenen Akzentuierung einstimmiger, nicht wiederholter Rhythmen. Diese Erkenntnisse lassen sich zu drei Akzentregeln zusammenfassen (vgl. Monahan et al., 1987, S. 577; Monahan, 1993, S. 126):

1. Ein (relativ zu den umgebenden Tönen) langer Ton bzw. ein Ton, der ein langes *Interonset-Intervall* (IOI) beginnt, bekommt den stärksten Akzent in einem zeitlichen Pattern.

2. Einen schwächeren Akzent bekommt ein Ton, der ein längeres IOI beendet, wenn dieser Ton der Anfang einer Tongruppe von drei oder mehr Tönen ist.
3. Der erste Ton einer Tonfolge bekommt einen Akzent.

Zur Vereinfachung der Akzentermittlung führt Monahan den Begriff der zeitlichen Kontur ein. Die zeitliche Kontur ist die Abfolge der Unterschiede der Tondauern. Dabei reichen drei Zeichen aus, um diese Abfolge zu charakterisieren: „+“ für eine gegenüber dem Vorgänger-IOI längere Dauer, „-“ für eine kürzere und „=“ für eine gleichlange Dauer.

Zumindest für einstimmige und nicht zu komplexe Tonfolgen lassen sich Monahans Akzentregeln so interpretieren, dass bei jeder Änderung der Tondauer (jedem „+“ und jedem „-“), der entsprechende Ton einen Akzent trägt, wobei Akzente mit „+“ stärker sind.

Ziel der Akzentbestimmung ist die Ermittlung der sogenannten ‚internen metrischen Uhr‘, d.h. eines Metrums, das bei einem Hörer durch die Akzente dieser Folge induziert wird. Die interne Uhr wird durch die beiden Größen Abstand zwischen den ‚Uhrschlägen‘ in Elementarpulsen und Phase bzw. Einsatz der Uhr relativ zum Beginn der rhythmischen Folge charakterisiert.

Nach Povel und Essens (1985) nennt Monahan (Monahan et al., 1987, S. 577-578) drei Regeln für einen Algorithmus, der die beste interne Uhr zu einer Folge von rhythmischen und mit Akzenten versehenen Ereignissen ermittelt:

- Die beste Uhr muss die Zahl der Elementarpulse des ganzen Musters ohne Rest teilen,
- die beste Uhr soll Schläge, die mit keinem *onset* der rhythmischen Folge zusammentreffen vermeiden,
- die beste Uhr soll Schläge, die mit *onsets* ohne natürlichen Akzent zusammentreffen, vermeiden.

In Ergänzung zu diesen Akzentregeln führt Monahan noch Ergebnisse von Vos (1977) an, die die wahrgenommene Gruppierung von Tönen zu Tongruppen aufgrund ihrer zeitlichen Eigenschaften beschreiben. Auch Vos bestätigt demnach, dass längere IOIs mit größerer Wahrscheinlichkeit als mit Akzent versehen wahrgenommen werden und rhythmische Gruppen beenden. Vos unterscheidet zusätzlich IOIs, in denen ein Ton zu hören ist und solche mit merklichem Pausenanteil: Ein langer klingender Ton wird wahrscheinlich als akzentuiert gehört, ein Ton, auf den eine längere Pause folgt als Gruppenende. Der stille Anteil eines IOIs sei damit für die Gruppierung zuständig, der klingende für den Akzent, wie es Monahan und Kollegen zusammenfassen (Monahan et al., 1987, S. 577). Monahan

selbst zieht eine Verbindung zur Gestaltheorie und im besonderen zum Gesetz der Nähe, das für den zeitlichen Abstand von Ereignissen ähnliche Voraussagen macht.

Aufgrund einer Studie von Clarke (1985), die den präzisen Umgang eines Konzertpianisten mit komplexen Rhythmen dokumentiert und analysiert, erweitert Monahan (1993) ihr Rhythmuskonzept, so dass es auch zur Beschreibung von mehrstimmiger, komplexer Musik dienen kann. Eine analytische Anwendung dieses Konzepts ist die präzisere Vorhersage der ‚internen metrischen Uhr(en)‘, die auch die rhythmische Information von mehrstimmigen Musikstücken strukturieren. Nach diesem Konzept besteht der psychologische Raum, in dem Rhythmen repräsentiert sind, aus drei Dimensionen:

- aus der metrischen Basis, die in der westlichen Musik fast ausschließlich binär oder ternär ist,
- aus der Form der Unterteilung der Beats, die gleich oder ungleich sein kann und bei der die Reihenfolge der ungleichen Beats eine Rolle spielt (eine längere Dauer markiert v.a. für Versuchspersonen mit geringer musikalischer Vorbildung einen Einsatz des Beats) und
- aus den hierarchischen Ebenen, auf denen sich rhythmische Phänomene abspielen und zu denen interne Uhren gesetzt werden können; das Setzen von mehreren internen Uhren setze jedoch größere musikalische Expertise voraus.

Da Monahan bis zu ihrem Tod leider keine Experimente zur Akzentuierung und den Gedächtnisleistungen für echte mehrstimmige Musik veröffentlicht hat, ist dieses mehrdimensionale Rhythmuskonzept nie auf seine Operationalisierbarkeit und seine Validität geprüft worden.

Für ihre Beschreibungen von einstimmigen kurzen Tonfolgen (Monahan & Carterette, 1985; Monahan et al., 1987) und echten Melodien (Monahan, 1993) reichen damit die oben angeführten beiden Akzentregeln bzw. der zusätzliche Abgleich mit den betonten Beats des zugrunde liegenden Metrums und die Gruppierungsregel nach Vos aus.

Tonhöhenakzente:

Für die Ermittlung von Akzenten, die durch den Tonhöhenverlauf einer Melodie oder einer kurzen Tonfolge zustande kommen, benutzt Monahan die von Thomassen (1982) empi-

risch gewonnenen Akzentregeln, die sich folgendermaßen zusammenfassen lassen (vgl. Monahan et al., 1987, S. 580):

- Bei Zwei-Ton-Motiven klingt der zweite Ton akzentuiert.
- Bei zwei aufeinanderfolgenden Intervallen in entgegengesetzter Richtung klingt der mittlere Ton akzentuiert.
- Bei drei oder vier aufeinanderfolgenden Tönen, die durch Intervalle getrennt sind, von denen eines deutlich größer ist als die anderen, bekommt der Ton nach diesem größeren Intervall einen Akzent.
- Wenn einem Ton ein Intervall von 4 oder mehr Halbtönen vorausgeht, so klingt er akzentuiert.
- Der Effekt der Konturumkehr (b) hat eine stärkere wahrgenommene Betonung zur Folge als der Effekt der relativen Intervallgröße (c und d).
- Der erste Ton einer Melodie/Tonfolge bekommt einen Akzent.

Die Folge der Tonhöhenakzente, die sich aus diesen Regeln ergeben, ist die einzige Repräsentation dieser Akzentdimension; d.h. in Monahans Modell wird nicht versucht, diese Tonhöhenakzente auf eine abstraktere Ebene zu reduzieren, wie sie für den Bereich der Tondauern die interne Uhr darstellt.

Die Gruppierung in der Dimension der Tonhöhe ist nach Monahan durch die Nähe der Töne hinsichtlich ihrer Tonhöhe (in Hertz) gegeben. Danach gilt eine Gruppe von Tönen als durch einen Ton abgeschlossen, wenn auf diesen Ton ein Sprung (oder ein wesentlich größeres Intervall gegenüber den Vorgängerintervallen) folgt.

Konsonanz der Akzentverteilungen und der Gruppierungen:

Mit Hilfe der Regeln für die Ermittlung der rhythmischen und der Tonhöhenakzente kann Monahan zu einer gegebenen Melodie zwei Akzentreihen ermitteln: die Folge der Tonhöhenakzente und die Schläge der ‚internen Uhr‘, die durch die rhythmischen Akzente induziert wird. Für die Beschreibung der Übereinstimmung definiert sie unter Verwendung der Terminologie und der theoretischen Überlegungen von Yeston (1976) drei prinzipielle Möglichkeiten: Die Konsonanz (das Zusammentreffen von Akzenten aus beiden Dimensionen), die phasenverschobene Konsonanz (das regelmäßige Abwechseln von rhythmischen und Tonhöhenakzenten) und die Dissonanz (das unvorhersehbare Nicht-Zusammentreffen von Akzenten aus beiden Dimensionen). Bedeutung – zumindest im Bereich der ‚norma-

len' westlichen Musik und der westlichen Hörer – für die herausgehobene Wahrnehmung und damit auch für die bevorzugte Enkodierung und Speicherung haben nach ihrer Vorstellung die Töne, bei denen Tonhöhenakzente auf Schläge der internen Uhr fallen, bei denen also Konsonanz zwischen den Dimensionen gegeben ist.

Die gleichen Möglichkeiten der Übereinstimmung bestehen auch für Gruppierungen, die innerhalb der beiden Dimensionen getrennt ermittelt werden. Stehen die rhythmische und die Gruppierung nach Tonhöhen in Konsonanz, so sollte dies für die Enkodierung und Speicherung der Töne in dieser Gruppe von Vorteil sein.

Monahans Modell der melodisch-rhythmischen Konsonanz zeigt eine große Nähe zu dem Modell von Boltz und Jones (1986), was Monahan und Kollegen selbst bemerken (Monahan et al., 1987, S. 581). Die Vorteile des eigenen Modells lägen jedoch in der größeren Differenziertheit: Bei Boltz und Jones werde die Gruppierung als Ebene von Übereinstimmung nicht beachtet und eine Beschreibung der Qualität der Übereinstimmung zwischen Akzentreihen – z.B. das Entdecken der phasenverschobenen Konsonanz – sei nach dem Modell von Boltz und Jones nicht möglich.

Das Modell von Boltz und Jones soll im folgenden Teilkapitel erörtert werden.

Empirische Prüfung des Modells:

Leider gibt es keine Veröffentlichung, in der Monahans Modell für die Vorhersage von Gedächtnisleistungen zu echten Melodien verwendet worden wäre. Eine systematische Untersuchung liegt nur für künstlich generierte kurze Tonfolgen vor (Monahan et al., 1987). Monahan und Kollegen verwendeten in dieser Studie ein Wiedererkennensparadigma, bei dem die Versuchspersonen die Identität von zwei nacheinander gehörten Tonsequenzen zu beurteilen hatten. Die beiden Sequenzen waren entweder identisch oder wichen in einem der neun Töne von einander ab. Nach Monahans Hypothese sollten die Abweichungen bei den Tönen bzw. bei den Sequenzen, bei denen die Akzente der beiden Dimensionen in Konsonanz stehen, am besten erkannt werden. Die Ergebnisse dieser sehr aufwendigen und komplexen Studie bestätigen insgesamt die Vorhersagen des Modells: Wenn zwischen Tonhöhenakzenten und rhythmischen Akzenten (interne Uhr) perfekte Übereinstimmung herrschte, waren die Wiedererkennungsleistungen signifikant am höchsten. Dies galt besonders für Töne, die einen rhythmischen Akzent aufgrund einer langen Tondauer erhielten. Stimmt die rhythmischen und die Tonhöhenakzente nicht

oder nur teilweise überein, so nahmen die Versuchspersonen Synkopen wahr und zeigten deutlich schlechtere Erkennensleistungen.

Zusätzlich zeigten sich noch einige weitere Zusammenhänge in den Ergebnissen, die für eine weitere Detaillierung des Modells hilfreich sind:

Wurde die Tonhöhenkontur durch die Abweichung in einem Ton geändert, so waren bessere Erkennensleistungen zu beobachten als bei Abweichungen ohne Konturänderungen.

Änderungen im Verlauf der melodischen Kontur hatten bessere Erkennensleistungen für den zweiten an der Konturänderung beteiligten Ton zur Folge.

Wurden die ersten drei Töne der Tonsequenz am Ende (als letzte Töne) wiederholt, so führte dies zu einer deutlich besseren Erkennensleistung. Die Versuchspersonen scheinen also Rhythmus und Tonhöhenfolge der Töne durch die Wiederholung gelernt zu haben.

Die musikalische Vorbildung der Versuchspersonen war von entscheidender Bedeutung für die Wiedererkennensleistungen. Dieser Faktor erklärte etwa 50% der Varianz in den Ergebnissen der Studie. Die Studierenden mit musikalischer Vorbildung zeigten deutlich bessere Leistungen als ihre Kommilitonen ohne solche.

Musikalisch vorgebildete und nicht-vorgebildete Versuchspersonen verarbeiteten und erinnerten die Tonsequenzen auf unterschiedliche Weise: Die Versuchspersonen ohne Vorbildung schienen weniger durch den induzierten metrischen Puls geleitet zu sein, ausschlaggebender für sie waren eher die langen gefüllten Töne. Auch die Dauernverhältnisse schienen für die nicht-vorgebildeten Versuchspersonen von geringer Bedeutung zu sein oder gar nicht erkannt zu werden. Sie schienen dagegen die Dauerninformationen nach dem einfachen ‚Länger-oder-Kürzer-Prinzip‘ (der zeitlichen Kontur) zu verarbeiten. Ganz ähnlich waren offensichtlich auch die hierarchischen Beziehungen der Töne hinsichtlich der Tonart der Sequenz für diese Gruppe von geringerer Bedeutung, stattdessen richtet sich die Orientierung vielmehr nach dem Abstand in der Tonhöhe, der für Gruppierungen verantwortlich war. Nur die musikalisch vorgebildeten Versuchspersonen schienen also die strukturellen Informationen der Sequenzen wie Dauernverhältnisse oder harmonische Verwandtschaften zwischen Tönen verarbeiten und speichern zu können. Dies deckt sich mit oben referierten Ergebnissen anderer Forschungsansätze, wie beispielsweise den Untersuchungen von Oura und Hatano (1988) oder Dowling et al. (1995). Ein Modell, was die Enkodierung und Speicherung von melodischem Material bei Laien beschreiben will, müsste also diese ‚Defizite in der strukturellen Verarbeitung‘ berücksichtigen.

Wie gesagt gelten diese Ergebnisse erst einmal nur für den Bereich der kurzen tonalen Tonfolgen. Für die echte Melodie *God save the queen*, die in veränderter metrischer und rhythmischer Gestalt präsentiert wurde, sodass Tonhöhen- und rhythmische Akzente nicht wie in der Originalgestalt zusammenfallen, führt Monahan nur das informelle Ergebnis an, dass nur ca. 30% der von ihr getesteten musikalisch Vorgebildeten und 10% der Nicht-Vorgebildeten diese Melodie erkannt hätten (Monahan, 1993, S.145).

Eine echte Prüfung der Gültigkeit und Vorhersagestärke ihres Modells für den Bereich der echten Melodien steht jedoch noch aus. Ist die zu behaltende Melodie in den Zusammenhang echter Musik eingebettet, z.B. Teil eines Popsongs, so ist zu erwarten, dass die Ermittlung der internen rhythmischen Uhr nur in geringem Maße durch die Dauern und Akzente der Melodie selbst bestimmt wird, sondern wahrscheinlich stärker auf den Beats des Songs insgesamt basiert, welche durch den Rhythmus des musikalischen Hintergrunds markiert werden. Dies dürfte die Akzentermittlung zumindest für den rhythmischen Bereich sehr vereinfachen, wirft aber gleichzeitig zusätzliche Fragen auf.

Beispielsweise ist die Bedeutung von Synkopen ungeklärt, die durch Betonungen der Melodiestimme auf Zählzeiten zustande kommen, die durch den musikalischen Hintergrund nicht betont sind. Werden Töne auf diesen Synkopen besser oder schlechter verarbeitet als Töne auf betonten Zählzeiten des Metrums des Songs?

Genauso ungeklärt ist, wie die Effekte von rhythmischer und Tonhöhenkonsonanz quantifiziert werden können, im Sinne des Kriteriums d) (Metrik), welches in der Einleitung zu diesem Kapitel genannt wurde. Es fehlt eine Metrik, nach der die Stärke der Einflüsse in Monahans Modell und deren Auswirkungen auf die Gedächtnisleistungen geordnet werden könnten.

Für die Adaption des Konsonanzmodells von Monahan für einen Bereich echter Musik bedarf es also weiterer Spezifizierungen und Präzisierungen.

Zusammenfassend lässt sich festhalten, dass das Konsonanzmodell von Monahan über ein ausgeklügeltes Repräsentationssystem verfügt, mit dem die natürlichen Akzente, die durch die Tonhöhenbewegung und die Folge der Tondauern zustande kommen, nachvollziehbar dargestellt werden können. Die Heuristik, die sie postuliert, ist die höhere Aufmerksamkeit und die priorisierte Enkodierung von Tönen, bei denen Akzente aus dem rhythmischen und dem Tonhöhenbereich zusammenfallen (konsonant sind). Die bessere Enkodierung dieser

Töne hat nach Monahans Annahme auch eine bessere Gedächtnisleistung für diese Töne zur Folge. Zum Nachweis ihres Modells liegen zwei methodisch verschiedene Studien vor. Die eine verwendet Ähnlichkeitsurteile von Versuchspersonen und die Methode der Multidimensionalen Skalierung zur Ermittlung der Dimensionen, die bei der Verarbeitung von kurzen Tonfolgen relevant sind (Monahan & Carterette, 1985). Die andere Studie prüft die Voraussagen des Konsonanzmodells mit Hilfe eines Wiedererkennenstests für kurze Tonfolgen (Monahan et al., 1987). In einer späteren Studie mussten Monahan und Kollegen jedoch feststellen, dass die Akzentstruktur zumindest kein universales Prinzip zur Vorhersage von Performanz-Leistungen in auditorischen Tests ist. Für das Entdecken von Abweichungen in Interonset-Intervallen erwies sich ein Modell, das auf dem Weberschen Verhältnisgesetz beruht, als besserer Prädiktor als das rhythmische Akzentmodell (Monahan & Hirsh, 1990; s.a. Hirsh et al., 1990). Zusätzlich fehlt Monahans Modell auch eine Metrik, nach der die Effekte von Einflussfaktoren und Effektgrößen quantifiziert werden können. Ein großer Nachteil für die praktische Verwendung ihres Modells ist schließlich der bislang fehlende Nachweis der Gültigkeit auch für den Bereich der echten Melodien bzw. von Melodien, die in echter Musik erklingen.

3.3.3.2 Die Joint Accent Hypothesis von Boltz und Jones

Die *Joint Accent Hypothesis*, die von Marilyn Boltz und Mari Riess Jones über mehrere Jahre und mehrere Veröffentlichungen hinweg ausgearbeitet und empirisch geprüft wurde, zeigt sowohl im konzeptuellen Ansatz wie auch in vielen Details der strukturellen Beschreibung von Melodien große Ähnlichkeit mit dem Modell der melodisch-rhythmischen Konsonanz von Caroline B. Monahan. Auch Boltz und Jones gehen davon aus, dass das gleichzeitige Auftreten von Akzenten, die durch den Tonhöhenverlauf generiert werden und solchen, die durch die Sequenz der Tondauern entstehen, für die Wahrnehmung, Verarbeitung und Erinnerung von Melodien und kurzen Tonfolgen von größter Bedeutung sind. Zusätzlich betrachtet vor allem Boltz in den jüngeren Veröffentlichungen auch harmonische Akzente. Im Gegensatz zu Monahan beschränken Boltz und Jones den Geltungsbereich ihres Modells nicht auf Erinnerung von Tonhöhenverläufen. Sie können auch allgemeine Vorhersagen ableiten zum Verlauf der menschlichen Aufmerksamkeit in zeitgebundenen Domänen sowie zu menschlichen Schätzungen und Reproduktionen von Zeitintervallen. So haben sich Jones und Boltz in den letzten Jahren z.T. unabhängig voneinander bemüht, das Modell, das anhand der Melodieverarbeitung und -erinnerung entwickelt wurde, zu einer allgemeinen Aufmerksamkeits- (Large & Jones, 1999; Jones et al., 2002),

Zeit-Wahrnehmungs- (Jones & Boltz, 1989; Yee, Holleran & Jones, 1994) und Zeit-Erinnerungstheorie (Boltz, 1992, 1995, 1998) weiterzuentwickeln. Im Folgenden wird jedoch nur die für das Melodiegedächtnis relevante *Joint Accent Hypothesis* und die strukturelle Beschreibung der dynamischen Form („dynamic shape“) als Teil des gesamten theoretischen Rahmens zusammengefasst.

Strukturelle Beschreibung von Melodien:

Wie bei Monahan betrachten Boltz und Jones Akzente, die in den Dimensionen des Tonhöhen- und des rhythmischen Dauernverlaufs sowie durch die Harmonik einer Melodie voneinander getrennt entstehen, um daraus eine strukturelle Beschreibung im Sinne von Akzentverteilungen abzuleiten.

Tonhöhenakzente:

Tonhöhenakzente können nach der Erklärung von Jones (1987, S. 623; 1993, S. 76-77) auf drei verschiedene Arten zustande kommen:

- a) In einer Tonsequenz bekommen die lokal höchsten oder tiefsten Töne einen Akzent. Dies entspricht der Akzentuierung bei Konturänderung, d.h., der Ton, bei dem eine Konturänderung stattfindet, bekommt einen Akzent. Dieses Kriterium entspricht exakt dem von Monahan formulierten Regeln für Konturumkehrungen.
- b) In einer Reihe von aufeinanderfolgenden Intervallen, bekommt derjenige Ton einen Akzent, dem das relativ größte Intervall vorausgeht (mindestens vier oder fünf Halbtöne). Dieses Kriterium entspricht fast genau dem von Monahan formulierten Kriterien für Intervallsprünge.
- c) Der erste Ton einer Melodie/Tonfolge bekommt einen Akzent.

Die Akzentermittlung anhand des Tonhöhenverlaufs von Boltz und Jones entspricht also den wichtigsten Regeln aus Monahans Modell. Zu beachten ist jedoch, dass die Ermittlung von Akzenten bei Konturumkehrungen in den sukzessiven Veröffentlichungen von Boltz und vor allem Jones nicht einheitlich gehandhabt wird. So bekommt nach der ausführlichen Beschreibung in Boltz & Jones (1986, S. 408) der Ton, der auf eine Konturumkehr folgt, einen Akzent, während in den späteren Veröffentlichungen der Ton der Konturumkehr selbst akzentuiert wird, wie an Beispielen verdeutlicht wird (Jones, 1987, S. 623; Jones & Ralston, 1991, S. 9-10; Jones, 1993, 76-77). Da die letztere Bestimmung von Konturumkehrungs-Akzenten im Lichte anderer Veröffentlichungen zum Thema plausibler er-

scheint und vor allem in den neueren Veröffentlichungen von Jones angeführt wird, soll sie im folgenden wie in Regel a) beschrieben angenommen werden.

Zeitliche Akzente:

Auch für die Ermittlung der zeitlichen oder rhythmischen Akzente geben Boltz und Jones Regeln an, die sich folgendermaßen zusammenfassen lassen:

- a) Jede sehr lange oder kurze Note im Vergleich mit dem Kontext der umgebenden Tondauern bekommt einen Akzent.
- b) Ein (relativ zum umgebenden Kontext) langer Ton, dem eine Pause folgt, bekommt einen Akzent.
- c) Ein Ton, dem eine Pause vorausgeht, bekommt einen Akzent, jedoch nur, wenn der Pause kein langer Ton vorausgeht, d.h. wenn Regel b) nicht greift.
- d) Der erste Ton einer Folge bekommt einen Akzent.

Von diesen vier Regeln entsprechen die Regeln a), c) und d) den von Monahan genannten und oben aufgeführten Regeln zu Ermittlung von rhythmischen Akzenten. Boltz und Jones geben zusätzlich auch noch sehr kurzen Tönen einen Akzent (Regel b)), was jedoch in den von ihnen konstruierten Experimentalmelodien nie zur Anwendung kommt. Wie Monahan bemerkt, unterscheiden Boltz und Jones jedoch nicht zwischen Akzentsetzung und Gruppierung, d.h. in ihrem Modell wird eine Melodie als fortlaufende Sequenz mit Akzenten in zwei verschiedenen Dimensionen repräsentiert, während Monahan noch mit Hilfe der Erkenntnisse von Vos (1977) Gruppierungen durch die Pausenverteilung bzw. durch die Pausenanteile in den Interonset-Intervallen festlegen kann.

Harmonische Akzente:

Die Definition eines harmonischen Akzentes, die Boltz in ihren Studien der 90er Jahre gibt, ist sehr einfach:

- Melodische Töne, die mit Grundton, Terz, oder Quinte der Tonika der Melodie (bzw. eines zugleich erklingenden Akkordes) identisch sind, bekommen einen Akzent.

Ganz ähnlich definiert Jones den „tonal end accent“, den sie in ihrer Systematik zu den melodischen Akzenten zählt (Jones, 1987, S. 623). Dieses Kriterium bezieht den harmonischen Kontext der jeweiligen Melodie ein und geht insofern über die Beschreibung von

Tonhöhenakzenten von Monahan hinaus. Allerdings vergibt Boltz in den entsprechenden Studien, in denen harmonische Akzente betrachtet werden, nur harmonische Akzente für Töne, die auf Phrasenenden zu hören sind. Jedoch bleibt unklar, wie Phrasenenden im ersten Schritt definiert werden, um dort die Übereinstimmung mit Akkordtönen zu identifizieren. Boltz scheint in den entsprechenden Studien (Boltz, 1991, 1993, 1998) Phrasenenden aufgrund von rhythmischen Informationen (längere Tondauern) sowie melodischen (Konturumkehr und/oder Sprung) und formalen Einschnitten (Wiederholungen) zu erschließen. Ohne explizites Kriterium für die Ermittlung von Phrasenenden dürfte aber die Formalisierung der Ermittlung harmonischer Akzente schwierig sein, wie sie beispielsweise für ein automatisches Melodieanalyse-System notwendig wäre. Zudem widerspricht das Betrachten von harmonischen Akzenten nur auf Phrasenenden der Idee einer unabhängigen Ermittlung und eines Vergleiches von Akzenten verschiedener musikalischer Dimensionen.

Gemeinsame Akzente, Ähnlichkeit und Verarbeitungsschwierigkeit:

Wie Monahan gehen Boltz und Jones davon aus, dass Töne die sowohl einen zeitlichen als auch einen melodischen oder harmonischen Akzent erhalten, stärker betont und insgesamt strukturell wichtiger sind. Nach der einfachen Darstellung, die Jones (1987, S. 625-626; 1993, S. 78-79) liefert, bekommt jede Note auf ihrem Onset den Akzentwert 1. Kommen ein melodischer und/oder zeitlicher Akzent bei dieser Note hinzu, so erhält sie den Akzentwert 2 bzw. 3. Aufbauend auf dieser simplen Akzentaddition definieren Boltz und Jones nun die Ähnlichkeit zwischen Melodien und die Verarbeitungsschwierigkeit einer Melodie, die auch für deren Erinnerbarkeit verantwortlich ist.

Für die Bestimmung der Ähnlichkeit führt Jones (Jones et al., 1987; Jones, 1993) das Konzept der dynamischen Form („dynamic shape“) ein. Die dynamische Form ist das relative *timing* der melodischen Akzente aus Konturumkehrungen in Bezug auf die zeitlichen Akzente. Nach diesem Konzept erinnern sich Versuchspersonen bei neuen Melodien vor allem an die dynamische Form, d.h. den relativen Zeitpunkt von Konturumkehrungen im Vergleich zu den zeitlichen Akzenten. Je mehr die dynamischen Formen zweier Melodien übereinstimmen, desto ähnlicher werden sie wahrgenommen und desto häufiger werden sie verwechselt. Allerdings gibt Jones keine Metrik an, nach der die Ähnlichkeit von Melodien zu quantisieren wäre. In den Experimenten mit ihren Kollegen werden daher auch nur künstliche Melodien verwendet, bei denen im vorhinein klar ist, dass sie hinsichtlich der

dynamischen Form identisch sind oder nicht. Für die Ähnlichkeitsbestimmung von Melodien in existierender Musik, wäre eine solche Metrik auf jeden Fall noch zu erarbeiten.

Im Gegensatz zur Ähnlichkeitsbestimmung geben Boltz und Jones für die Ermittlung der Verarbeitungsschwierigkeit eine Metrik an, die auch für Melodien in echter Musik anwendbar ist. So wird ihre Joint Accent Hypothesis in diesem Bereich überprüfbar. Auch diese Metrik geht von der Übereinstimmung von Akzenten in melodischer und zeitlicher Dimension aus bzw. von der Übereinstimmung in den Abständen zwischen melodischen und zeitlichen Akzenten. Wie Jones an vielen Beispielen verdeutlicht (z.B. Jones, 1987, S. 629-632), lässt sich melodische Einfachheit durch zwei Quotienten ausdrücken:

Zum einen beschreibt das Verhältnis zwischen Akzenten mit dem Akzentwert 3 und Akzenten mit dem Akzentwert 2 die Einfachheit einer Melodie: Dabei stehen kleine ganzzahlige Quotienten für einfache Melodien und komplexere Verhältnisse für schwierigere Melodien. Zum anderen beschreibt der Quotient der Abstände zwischen melodischen Akzenten und zeitlichen Akzenten die Schwierigkeit einer Melodie. Auch hier definieren kleine ganzzahlige Quotienten einfache Melodien.

Für diese Übereinstimmungen zwischen Akzentabständen geben Boltz und Jones (1986) eine Metrik an, die durch das Indexmaß A wiedergegeben wird, von Jones aber später als „probably premature“ bezeichnet wird (Jones, 1987, S: 632). Da das Indexmaß A sich jedoch in der Studie von 1986 als guter Prädiktor für die Einfachheit einer Melodie und für das Behalten im Gedächtnis erwies und die Quotienten der Akzentabstände relativ gut in ein skalierbares Maß zusammenfasst, soll die Bestimmung von A hier kurz erläutert werden.

Die Metrik lässt sich mit der folgenden Gleichung beschreiben:

$$A = T + U$$

Dabei steht A für den Index der Verarbeitungsschwierigkeit. T ist die Anzahl der Unregelmäßigkeiten in der Tonhöhenfolge (gemeint sind im speziellen die Sprünge) und U ist die Summe der unvorhersehbaren Tonhöhenakzente. Das Indexmaß setzt sich also zusammen aus der Zahl der melodischen Akzente (durch Sprung erzeugt) und der Zahl der (nach der Verteilung der Dauernakzente) unerwarteten melodischen Akzente.

Zur Ermittlung von T werden alle Sprünge (Intervalle > 2 Halbtöne) gezählt und dazu noch ein Akzent für den ersten Ton einer Folge addiert.

U ist die Anzahl der unerwarteten Akzente. U ist die Summe aus U_p (unerwartete Akzente durch Sprung) und U_c (unerwartete Akzente durch Konturumkehr). U wird in drei Schritten ermittelt:

1. Die zeitlichen Akzente einer Melodie werden nach den vier oben genannten Regeln ermittelt und die Periodenlänge zwischen den zeitlichen Akzenten in Beats bestimmt. Voraussetzung ist dabei, dass sich eine regelmäßige Periode mit einheitlicher Länge zwischen den zeitlichen Akzenten finden lässt.
2. Die melodischen Akzente werden nach den oben genannten vier (bzw. drei: a), b), d)) Regeln ermittelt und die Abstände (Perioden) zwischen den Akzenten aufgrund von Sprüngen und von Konturumkehrungen getrennt in Beats gezählt.
- 3.a Alle Perioden (Abstände) zwischen Akzenten, die aufgrund von Sprüngen zustande kommen und die nicht die gleiche Länge in Beats haben wie die Perioden der zeitlichen Akzente, werden addiert (U_p).
- 3.b Alle Perioden (Abstände) zwischen Akzenten, die aufgrund von Konturumkehr zustande kommen und die nicht die gleiche Länge in Beats haben wie die Perioden der zeitlichen Akzente, werden addiert (U_c).

Dann lässt sich A bestimmen nach:

$$A = T + (U_p + U_c)$$

Die Schwierigkeit bei der Verarbeitung und Wiedergabe einer Melodie steigt also mit dem Index A . Obwohl A relativ präzise bei Boltz und Jones (1986, S. 407-412) definiert ist, bleiben dennoch einige Fragen hinsichtlich der Konstruktion dieses Schwierigkeitsmaßes offen:

Im Gegensatz zu Monahan geben Boltz und Jones nur indirekte Hinweise für eine theoretische oder empirische Fundierung der von ihnen verwendeten Akzentregeln (z.B. Thomasen, 1982; Lerdahl & Jackendoff &, 1983). Noch weniger wird die Konstruktion von A begründet oder diskutiert. Beispielsweise bleibt unbeantwortet, warum nur die Anzahl der Sprünge als T gezählt wird und nicht auch die Konturumkehrungen, wo doch auch diese eine Änderung des melodischen Verlaufs darstellen. Keine Angaben machen die beiden Forscherinnen darüber, wie unregelmäßige Perioden zwischen zeitlichen Akzenten behandelt werden sollen. Für existierende Musik, die einen Grundbeat enthält, lässt sich jedoch davon ausgehen, dass dieser (regelmäßige) Beat für die Ermittlung der zeitlichen Perioden

herangezogen werden kann. Dieser Grundbeat ist zumindest in populärer Musik in der Regel durch regelmäßige metrische Akzente gekennzeichnet.

Empirische Prüfung des Modells:

Die Ähnlichkeit von kurzen Melodien bzw. die Tendenz bestimmte Melodien zu verwechseln, untersuchte Jones mit Kollegen (Jones et al., 1987) und konnte die Ergebnisse dieser Studie einige Jahre später replizieren (Jones & Ralston, 1991). In diesen Untersuchungen mit wenig musikalisch vorgebildeten Studenten konnte sie zeigen, dass tatsächlich Melodien, die die gleiche dynamische Form besaßen, viel häufiger miteinander verwechselt wurden als mit Melodien, die eine andere dynamische Form besaßen, ansonsten jedoch gleich konstruiert waren. Dies galt besonders dann, wenn die Versuchspersonen die Melodien nur wenige Male gehört hatten (dreimal). Auch bei sieben- bzw. elfmaligem Hören verwechselten die Versuchspersonen neue Melodien mit gleicher dynamischer Form mit den Originalmelodien, wenn auch in geringerem Maße. Jones nimmt deshalb an, dass die Versuchspersonen beim Hören einer Melodie eine Gedächtnisrepräsentation aufbauen, die in etwa der dynamischen Form entspricht: „This suggests that people attend to and remember the dynamic shape of a tune ...“ (Jones, 1993, S.84). Auf diese globale Repräsentation einer Melodie als dynamischer Form verlassen sich die Versuchspersonen anscheinend um so mehr, je weniger sie mit der Melodie bekannt sind, d.h. je weniger Details sie aufnehmen konnten: „[...] less experience with a tune would increase the chance that a listener would remember it globally, in terms of a rhythmical contour, rather than precisely, in terms of fixed pitch intervals.“ (Jones, 1987, S. 627-628).

Von grundlegender Bedeutung erwies sich der Rhythmus für die Verwechslungen im Gedächtnis (Jones & Ralston, 1991). Wurden die schon bekannten Melodien in einem neuen Rhythmus präsentiert, so wurden sie sehr häufig als neu bezeichnet, obwohl die Versuchspersonen ausdrücklich instruiert waren, den Rhythmus in allen Phasen des Experiments zu ignorieren. Der Rhythmus scheint die Verabreichung und die Gedächtnisrepräsentation einer Melodie grundlegend zu beeinflussen, auch wenn er nicht bewusst beachtet wird. Die Gedächtnisrepräsentation einer dynamischen Form besteht also nicht nur aus der relativen Lage der melodischen und zeitlichen Akzente, sondern bezieht auch den unterliegenden Rhythmus mit ein.

Wie gut das Indexmaß A die Schwierigkeit von Melodien für die Verarbeitung und das Behalten vorhersagt, wurde bislang nur in einer Studie evaluiert (Boltz & Jones, 1986). Danach ergab sich für den Zusammenhang zwischen A und den Fehlern bei der Wiedergabe der Versuchspersonen ein Korrelationskoeffizient (Produkt-Moment) von 0,85; ein relativ starker Zusammenhang im Vergleich mit einem rekursiven Regelmodell für die Melodieverarbeitung, das Boltz und Jones gleichzeitig in seiner Vorhersagestärke für Wiedergabefehler testeten. Das Maß dieses Modells korrelierte nur zu 0,71 mit den Fehlern der Versuchspersonen. Allerdings ist bei dieser Studie zu beachten, dass die Wiedergabe der Melodien als Melodiediktat erfolgte und insofern nur Musikstudenten als Versuchspersonen rekrutiert wurden. Aus diesem Grund ist die in diesem Versuch gefundene Vorhersagestärke von A für ‚normale‘ Versuchspersonen bisher nicht generalisierbar.

Die Bedeutung der harmonischen Akzente auf Phrasenenden untersuchte Boltz in einer Reihe von Studien in den 90er Jahren (z.B. Boltz, 1991, 1993, 1998). Die empirischen Ergebnisse der Studien bestätigen insgesamt die Joint Accent Hypothesis:

Fallen Dreiklangstöne auf Phrasenenden mit rhythmisch-zeitlichen Akzenten zusammen, so werden diese sehr gut erinnert. Als häufiger und musikalisch sehr nachvollziehbarer Fehler zeigte sich auf diesen Phrasenenden eine Verwechslung mit anderen Dreiklangstönen. Es muss jedoch ein zeitlich-rhythmischer Akzent die Dreiklangstöne unterstreichen, damit diese auf Phrasenenden effektiv kodiert werden können. Fiel dieser Akzent nicht mit dem harmonischen zusammen, so zeigte sich keine bessere Erinnerung an die Dreiklangstöne. Zusätzlich zu diesen erwarteten Effekten konnte Boltz (1991) noch einen starken Einfluß der melodischen Struktur der einzelnen Melodien beobachten. So spielten die Anzahl der melodischen Phrasen, die Anzahl der Konturänderungen (besonders, wenn sie nicht mit Anfang oder Ende einer Phrase zusammenfielen) sowie die Interaktion zwischen melodischen und rhythmischen Akzenten für die Behaltensleistung eine Rolle. Erstaunlich ist, dass Boltz diese Effekte, die ja durch die Joint Accent Hypothesis so oder so ähnlich vorhergesagt werden, nicht durch die Einbeziehung der dynamischen Form bzw. durch den Index für die Verarbeitungsschwierigkeit A kontrolliert hat. Durch die Aufnahme dieser Faktoren in das Modell hätte sich sicherlich ein noch größerer Teil der Varianz in den experimentellen Ergebnissen aufgeklärt werden können. Auch die Generalisierbarkeit der Ergebnisse von Boltz (1991) ist fraglich, da das Melodiediktat mit Musikstudenten als experimentelle Methode verwendet wurde.

Jedoch kommen die Folgestudien (Boltz, 1993, 1998, 1999), die die Wiedererkennensleistungen von Studenten mit geringer und mittlerer musikalischer Vorbildung bzw. die Genauigkeit der Reproduktion der Dauern von Melodien untersuchen, zu ähnlichen Ergebnissen, die die Joint Accent Hypothesis im Hinblick auf zeitliche und harmonische Akzente bestätigen. In den 1998 und 1999 veröffentlichten Studien kann Boltz nicht nur zeigen, dass die Erinnerung an die Melodiedauer und die strukturellen Merkmale von Rhythmus und Tonhöhe bei Melodien mit gemeinsamen Akzenten besser ist. Auch wurden die melodische und die zeitlich-rhythmische Dimension dieser akzentkohärenten Melodien gemeinsam enkodiert, und zwar auch dann, wenn die Versuchsinstruktion die Enkodierung nur einer Dimension vorgab. Das bedeutet nach Boltz' Interpretation, dass sich Melodien mit einer geeigneten Struktur (gemeinsame Akzente) immer als Ganzes ins Gedächtnis einprägen, auch wenn Tonhöhe oder Rhythmus subjektiv nicht beachtet werden. Bei Melodien, die einer gemeinsamen Akzentstruktur entbehren, findet diese automatische Enkodierung aller Dimensionen nicht statt. Von diesen Melodien wird von den Versuchspersonen vorrangig nur das erinnert, was subjektiv auch beachtet wurde (Rhythmus vs. Tonhöhenverlauf). Diese Idee beschreibt Boltz mit dem Stichwort „structural remembering approach“. Danach hängt der Modus der Enkodierung (gesamtes Objekt vs. Dimension eines Objektes) zum einen von der Struktur des zu Erinnernden Objektes (z.B. Melodie) ab und zum anderen von dem Grad der Bekanntheit bzw. des Erlernens des Objektes. Erst nach mindestens viermaligem Hören konnten die Versuchspersonen mit geringer musikalischer Vorbildung offenbar die akzentkohärente Struktur der Testmelodien erfassen und so die Melodie als Ganzes enkodieren und erinnern (Boltz, 1999).

Verwunderlich ist bei den Veröffentlichungen von 1998 und 1999, dass Boltz die dynamische Form und den Schwierigkeitsindex A nicht zur Analyse der Versuchsergebnisse heranzieht, sondern nur ganz allgemein zwischen akzentkohärenten und willkürlich geschaffenen nicht-kohärenten Melodien unterscheidet.

Eine weitere Richtung, in die Boltz (1993) und Jones (Jones et al., 2002) die Joint Accent Hypothesis und die Idee der gleichen Perioden zwischen Akzenten erweitern, ist die Generierung von Erwartungen. Nach Boltz (1993) erzeugen rhythmische Akzente, die jeweils den gleichen Abstand voneinander haben (gleiche Periodenlänge) und mit harmonischen Akzenten zusammenfallen, die Erwartung von weiteren rhythmisch akzentuierten Dreiklangstönen im weiteren Verlauf der Melodie. Wenn diese Erwartungen durch bestimmte Töne erfüllt werden, so prägen sich diese Töne dem Gedächtnis besonders gut ein und die

Erinnerungsleistung für diese Töne ist besser als wenn keine regelhafte Akzentstruktur gegeben ist. Voraussetzung ist, dass durch gleiche Akzentperioden eine erhöhte Aufmerksamkeit auf ein Ereignis zu einem erwarteten Zeitpunkt gelenkt wird. Die Gedächtnisleistungen – auch für Einzeltöne, wie Jones und Mitarbeiter (2002) zeigten – sind für Ereignisse auf die erwarteten Zeitpunkten besonders gut. Dieser Ansatz lehnt sich an die Theorien und Modelle von Leonard Meyer (z.B. Meyer, 1956; Rosner & Meyer, 1986) und Eugene Narmour (z.B. 1989) an und wurde im Besonderen von Mark Schmuckler weiterentwickelt (z.B. Schmuckler & Boltz, 1994; Schmuckler, 1997). Dieser Erwartungsansatz soll deshalb in der von Schmuckler ausgearbeiteten Form im nächsten Unterkapitel ausführlich vorgestellt werden.

Zusammenfassung und Kritik:

Die Joint Accent Hypothesis geht – ähnlich wie das rhythmisch-melodische Konsonanzmodell von Monahan – von der These aus, dass Töne, bei denen Akzente aus den Dimensionen Tonhöhenverlauf, Rhythmus und Harmonik zusammentreffen, für die Identität einer Melodie strukturell wichtiger sind als Töne, die keine gemeinsamen Akzente erhalten.

Auch die Akzentermittlung geschieht bei Boltz und Jones nach fast den gleichen Regeln wie bei Monahan. Zusätzlich beziehen Boltz und Jones jedoch auch Akzente aus der harmonischen Dimension mit ein.

Drei Arten von Voraussagen innerhalb des Modells von Boltz und Jones werden durch die strukturelle Beschreibung von Melodien als Sequenzen von Akzentverteilungen möglich:

1. Melodien, bei denen die relative Verteilung von melodischen und rhythmisch-zeitlichen Akzenten identisch oder ähnlich ist, werden häufig miteinander verwechselt. Die Repräsentation der relativen Verteilungen von melodischen und rhythmischen Akzenten nennen Boltz und Jones dynamische Form („dynamic shape“). Die Ähnlichkeit in der dynamischen Form kann somit als Maß für die kognitive Ähnlichkeit von Melodien genommen werden.
2. Melodien, bei denen die Akzente in der melodischen und in der rhythmischen Dimension regelmäßige und gleiche Abstände haben, sind leichter zu enkodieren, zu speichern und wiederzugeben als solche mit unregelmäßigen Akzentabständen. Es lässt sich das Maß A für die Verarbeitungsschwierigkeit von Melodien auf Grund der Akzentstruktur konstruieren, welches einen Vorhersagewert für die Anzahl der Fehler bei der Melodiewiedergabe aus dem Gedächtnis hat. Speziell für

Töne, bei denen rhythmische (Phrasenenden) und harmonische Akzenten zusammenfallen, gilt, dass diese Töne besonders gut erinnert werden. Insgesamt werden Melodien, die regelmäßige Akzentabstände in mehreren Dimensionen aufweisen und damit einen niedrigen Wert von A haben, als Ganzes enkodiert und gespeichert, auch wenn subjektiv nur eine der Dimensionen (z.B. Rhythmus) beachtet wurde. Hat die Melodie eine unregelmäßige Akzentverteilung, so kann nur die beachtete Dimension enkodiert werden. Diesen Zusammenhang zwischen ‚guter‘ Struktur und mehrdimensionaler Enkodier- und Erinnerungsleistung nennt Boltz „structural remembering approach“.

3. Regelmäßige rhythmische Akzente veranlassen den Hörer, die einheitliche zeitliche Akzentperiode in die Zukunft zu extrapolieren und die vorhergesehenen Zeitpunkte mit höherer Aufmerksamkeit zu beachten. Fallen auf die rhythmischen Akzente Dreiklangstöne, so wird eine Erwartung für weitere Dreiklangstöne auf den zukünftigen vorhersehbaren rhythmischen Akzentpunkten generiert.

Die Joint Accent Hypothesis von Boltz und Jones weist mehrere Vorteile auf, die sie für eine formalisierte Analyse von Melodien und eine Vorhersage von Gedächtnisleistungen gut brauchbar erscheinen lassen. Zu diesen Vorteilen gehört die Betrachtung der relativen Lage von rhythmischen und harmonischen Akzenten zueinander, bei der Phasenverschiebungen, die nur wenig an der Stabilität eines Strukturmusters ändern, nur gering gewichtet oder ‚bestraft‘ werden. Insgesamt gesehen birgt die betrachtete Akzentstruktur, der *dynamic shape*, eine konzeptionelle Ähnlichkeit zu der Melodiestructur, die nach Sloboda und Parker (1985; s.o.) zuerst beim Hören von unbekannten Melodien verarbeitet und im Gedächtnis gespeichert wird. Auch intuitiv und im Lichte früherer Forschungen zur Ähnlichkeit von Melodien klingt die Joint Accent Hypothesis plausibel. So ist beispielsweise eines der von Steinbeck zur Ermittlung der Melodieähnlichkeit herangezogenen Merkmale die Tonhöhe auf den Taktschwerpunkten (z.B. Steinbeck, 1982: Merkmal 22, S. 161). In die gleiche Richtung gehen auch die Vergleiche von Tönen auf betonten Taktzeiten, die Leppig in seinen Ähnlichkeitsuntersuchungen „Betonungsaffinitäten“ nennt (Leppig, 1987a, S. 146-147).

Von großem Vorteil ist bei der Joint Accent Hypothesis die Angabe einer exakten Metrik zumindest für den Bereich der Verarbeitungsschwierigkeit, die eine empirische Überprüfung ermöglicht. Seltsam erscheint jedoch, dass das Indexmaß A dieser Metrik mit Ausnahme einer Veröffentlichung (Boltz & Jones, 1986) nicht weiter für die Analyse von Versuchsergebnissen verwendet wurde. Jones selbst hält das Indexmaß für die Studie von

1986 zwar für angemessen, wendet sich aber dagegen, es als globales Maß für die Einfachheit von Melodien zu verwenden (Jones, 1987, S. 632). Eher sollten je nach experimenteller Aufgabe neue Ähnlichkeitsmaße entwickelt werden, die ebenso auf der Basis von Akzentabständen operieren.

Für den Aspekt der kognitiven Ähnlichkeit, die durch eine Ähnlichkeit in der dynamischen Form bedingt ist, schlagen Boltz und Jones leider kein Maß vor, sodass die Analyse der Ähnlichkeit zwischen echten Melodien mit der Konstruktion einer Metrik für die dynamische Form beginnen müsste.

Von Nachteil für die Verwendung des Modells von Boltz ist die Tatsache, dass es bisher weder mit mehrstimmiger Musik getestet wurde noch mit Musik, bei der ein rhythmisches Akzentmuster durch die Begleitung vorgegeben wird. Die Brauchbarkeit des Modells für diese Bereiche ist erst noch nachzuweisen.

Schließlich ist auch unklar, wie und ob sich die Beschreibung der dynamischen Form auch auf niedrigeren Ebenen des musikalischen Geschehens, z.B. auf dem Niveau von einzelnen Motiven anwenden lässt. Sind charakteristische Motive von wenigen Tönen, die i.d.R. als Bausteine für größere melodische Zusammenhänge dienen, auch in ihrer Ähnlichkeit und Verarbeitungsschwierigkeit durch die dynamische Form und das Maß A zu erfassen? Eine operative Formulierung dieser Frage und eine empirische Antwort stehen bislang aus.

Der Unterschied der Joint Accent Hypothesis zum melodisch-rhythmischen Konsonanzmodell von Monahan ist insgesamt gesehen nicht besonders groß. Die Unterschiede, die die Autoren jeweils selbst anführen (Jones: Unterscheidung globaler und lokaler Konturakzent; Monahan: Gruppierung durch rhythmische Parameter) haben innerhalb der jeweiligen Modelle nur Detailbedeutung und sind für die Erklärung der experimentellen Ergebnisse nur von sehr untergeordneter Bedeutung. Letztlich gehen die Joint Accent Hypothesis und das melodisch-rhythmische Konsonanzmodell von derselben Idee aus und beschreiben die zu untersuchenden Melodien sehr ähnlich.

Dass der Ansatz, Melodien in Hinblick auf gemeinsame Akzentstrukturen zu untersuchen, zur Analyse für eine Reihe von Verarbeitungs- und Produktionsvorgängen erfolgreich eingesetzt werden kann, zeigen einige Nachfolgestudien, die auf die Idee und die Akzentermittlungsverfahren von Monahan und Boltz und Jones zurückgreifen. Zu den in diesem Zusammenhang wichtigen Studien gehören die von Smith & Cuddy (1989) und Drake et al. (1991).

3.3.3.3 Das Konzept der musikalischen Erwartungen von Mark A. Schmuckler

Stärker als Caroline Monahan und Marilyn Boltz und Mari Riess Jones lehnt sich Mark A. Schmuckler bei der Entwicklung seines Konzeptes der musikalischen Erwartungen an bereits bestehende Theorie- und Modellansätze an. Wie er selbst ausführlich darlegt, stützt sich sein Modell der musikalischen Erwartung zum einen auf bereits vorhandene Erwartungskonzepte von Leonard Meyer (Meyer, 1956; Rosner & Meyer, 1982, 1986), Mari Riess Jones (Jones, 1981, 1982) und James C. Carlsen (Carlsen, 1981, 1982; Unyk & Carlsen, 1987).

Zum anderen verwendet Schmuckler Ergebnisse aus der kognitiven Musikpsychologie aus den Bereichen Gedächtnis (Bharucha & Stoeckig, 1987) und Strukturverarbeitung, wo er besonders auf das Implication-Realization-Modell von Eugene Narmour zurückgreift (Narmour, 1989, 1990).

Insofern ist sein Erwartungskonzept eine Zusammenführung und Operationalisierung von bereits existenten Modellen, die Schmuckler in einen Kontext stellt und empirisch prüft. Interessant im Zusammenhang dieser Arbeit ist vor allem der große Einfluß von musikalischen Erwartungen auf das Gedächtnis für musikalische Strukturen, welches Schmuckler in seinen jüngeren Arbeiten aufgezeigt hat (Chiappe & Schmuckler, 1997; Schmuckler, 1997). Die Darstellung seines Erwartungskonzeptes soll in zwei Schritten geschehen: Zuerst soll nachgezeichnet werden, welche musikalischen Struktureigenschaften Erwartungen generieren, um im zweiten Schritt Schmucklers Belege für den Zusammenhang zwischen Erwartungen und Gedächtnisleistungen anzuführen.

Generierung musikalischer Erwartungen:

Das Konzept der Erwartung definiert Schmuckler folgendermaßen:

Simply defined, „expectation“ refers to the anticipation of upcoming information based on past and current information. (Schmuckler, 1997, S. 292)

Schmuckler geht davon aus, dass es für den Bereich der westlichen tonalen Musik möglich ist, die Faktoren aus der gegenwärtigen und vergangenen musikalischen Information des Hörverlaufs zu ermitteln, die für Generierung von Erwartungen bei Hörern verantwortlich sind. Dabei lassen sich die Dimensionen der Erwartungsstärke und der Erwartungsspezifität quantisieren und mit der Stärke der Einflußfaktoren korrelieren. Mit Hilfe der so bestimmten objektiven Erwartungsfaktoren sind dann subjektive Einschätzungen der Erwartungen über eine musikalische Passage genauso wie Art und Ausmaß der Kodierung und

Verarbeitung dieser Passage vorhersagbar, was wiederum Schlüsse über das Gedächtnis und eine mögliche nachfolgende Produktion (Wiedergabe) dieser Passage ermöglicht.

Die Einflußfaktoren, die Schmuckler zur Bestimmung der Erwartung musikalischer Passagen verwendet sind primär harmonischer und melodischer Natur (vgl. Schmuckler, 1989). Schmuckler ergänzt in einer späteren Studie noch den Faktor Rhythmus (von Akkordfortschreitungen; Schmuckler & Boltz, 1994), welcher auch zur Generierung von Erwartungen beitragen kann.

Wie sehr ein Akkord erwartet wird, hängt nach seinem Modell v.a. von dem direkt vorangehenden Akkord ab und der Häufigkeit, mit der die beiden Akkorde in direkter Folge in Musik (eines bestimmten Stils) überhaupt erklingen. Diese Abfolgehäufigkeit wird von Hörern unbewusst gespeichert und steht nach Schmuckler in direkter Beziehung zu der Erwartungsstärke und -spezifität, die für eine harmonische Fortschreitung aufgrund eines bestimmten Akkordes generiert wird.

Als Quelle für die Häufigkeit von Abfolgen von zwei Akkorden verwendet Schmuckler die Tabelle von Akkordverbindungen von Piston (1978), die auf Auszählungen von Akkordfolgen in westlicher Kunstmusik beruht. Diese Tabelle ordnet einem Akkord auf einer der sieben Stufen der diatonischen Tonleiter die diatonischen Akkorde zu, die auf diesen Akkord oft, manchmal, selten oder nie folgen. Es werden dabei Ketten von nur zwei Akkorden betrachtet. Die Tabelle gibt also keine Auskunft über die Häufigkeit von längeren Akkordketten. Pistons Tabelle birgt durch die Angabe der abgestuften Häufigkeiten für den Übergang von einem Akkord zum nächsten eine große Ähnlichkeit zu Markov-Ketten 1. Ordnung, wie sie beispielsweise Steinbeck (1982, S. 198ff) beschreibt.

Schmuckler setzt diese Häufigkeitsklassen mit empirisch erhobenen Bewertungen der Erwartung von bestimmten Akkorden in Beziehung (Schmuckler, 1989, Exp. 2). Nach den Ergebnissen dieses Experimentes haben die von Piston ermittelten Häufigkeiten einen signifikanten Einfluß auf die harmonischen Erwartungen der Versuchspersonen. Allerdings richten sich die Versuchspersonen stärker nach dem harmonischen Kontext des verwendeten Experimentalstimulus (Schumann-Lied „Du Ring an meinem Finger“) als nach den Pistonschen Häufigkeiten. Denn auch bei Akkordverbindungen des Schumann-Liedes, die Piston als manchmal bzw. selten klassifiziert, erwarteten die Versuchspersonen am stärksten den tatsächlichen Akkord und nicht den nach Piston häufigsten. Hier zeigt sich auch nach Schmucklers Ansicht, dass die Akkordverbindungstabelle von Piston auf längere Akkordfolgen ausgedehnt oder anderweitig den harmonischen Kontext eines konkreten Sti-

ckes berücksichtigen müsste. Die deutlichsten Effekte erhielt Schmuckler für die Akkordverbindungen, die nach Piston als häufig gelten. Er geht deshalb davon aus, dass die Einteilung Pistons in den übrigen Klassen ‚manchmal‘, ‚selten‘ und ‚nie‘ strukturell nicht ganz richtig ist (es existieren mehr oder weniger als drei nicht-häufige Klassen) oder die Akkorde selbst z.T. falsch zugeordnet sind.

Darüber hinaus findet Schmuckler in diesem wie in den anderen Experimenten der selben Reihe eine hohe empirische Korrelation zwischen Erwartungsstärke und Spezifität, die ihn veranlasst, die beiden Dimensionen gemeinsam zu behandeln. Denn nach diesen Ergebnissen liegt die Vermutung nahe, dass die Versuchspersonen entweder nicht zwischen Erwartungsstärke und -spezifität unterscheiden können oder die eine Größe die andere in hohem Maße mitbedingt.

In einem späteren Versuch (Schmuckler & Boltz, 1994) verwenden Schmuckler und Boltz neben der Piston-Tabelle noch die im Experiment erhobenen empirischen Ergebnissen zur Beziehungsstärke von Akkorden in einem tonalen Kontext von Bharucha und Krumhansl (1983), die später in die Modellierung der neuronalen Architektur von tonalen und harmonischen Beziehungen von Bharucha und Kollegen einging (z.B. Bharucha, 1987, 1989). Dabei gibt Schmuckler jedoch weder eine Metrik noch einen Algorithmus an, nach dem die Stärke tonaler Beziehungen zwischen beliebigen Akkorden in existierender Musik mess- oder ermittelbar wäre. Auch in den Experimenten dieser Veröffentlichung zeigt sich ein deutlicher Einfluß der harmonischen Beziehungsstärke auf die Erwartungsbewertungen der Versuchspersonen. Da Boltz und Schmuckler außer der Beziehungsstärke der Akkorde auch noch zeitliche Variablen wie die Regelmäßigkeit der Akkordonsätze und Dauer des letzten Akkordes betrachten, finden sie eine deutliche Interaktion zwischen der harmonischen Beziehungsstärke und diesen zeitlichen Variablen. Diese Interaktion deutet nach der Interpretation der Autoren daraufhin, dass harmonische und rhythmische Erwartungen nicht unabhängig voneinander entstehen, sondern sich gegenseitig beeinflussen. Ein Strukturmodell, welches die durch den musikalischen Verlauf generierten Erwartungen beschreibt, muss also Dimensionen von Harmonik und Rhythmus einbeziehen und gegeneinander gewichten, wie dies z.B. in der Joint Accent Hypothesis von Jones und Boltz der Fall ist. Dass die Kompatibilität von harmonischem Rhythmus und metrischen Akzenten jedoch nicht automatisch zu besseren Verarbeitungs- und Behaltensleistungen führt, zeigt die Studie von Smith und Cuddy (1989). In der dort dokumentierten Versuchsreihe konnten die Autoren keine allgemeinen positiven Effekt von kompatiblen harmonischen und

metrischen Akzenten (bzw. Gruppierungen) finden. Nach der Interpretation der Autoren spielt die Konstruktion der Versuchsstimuli selbst eine große Rolle für die Effektivität der gemeinsamen Akzente (Smith & Cuddy, 1989, S. 467).

Für die Vorhersage der melodischen Erwartungen verwendet Schmuckler das bereits seit langem etablierte Erwartungskonzept von Meyer (Meyer, 1956, 1973), die Bestimmung der Tonalitätszugehörigkeit von melodischen Tönen nach Krumhansl (Krumhansl & Kessler, 1982) und empirischen Erkenntnisse über die Bedeutung der melodischen Kontur (v.a. von Dowling, s.o.).

Das Modell von Meyer erfuhr in den 80er Jahren einige empirische Überprüfungen und Belege (Rosner & Meyer, 1982, 1986). Nach diesem Modell bestehen Melodien in westlicher Musik vorrangig aus etwa einem halben Dutzend melodischer Schemata oder Prototypen, die als Struktur für kurze melodische Passagen dienen und deren Kontur beschreiben. Die Annahme von melodischen Prototypen für Motive und kurze Melodieabschnitte ist in der kognitiven Musikpsychologie weit verbreitet und findet sich beispielsweise auch bei Bigand (1993) als Theorieansatz. Von den melodischen Prototypen Meyers braucht Schmuckler nur zwei – das gap-fill Muster und das lineare Muster –, um den melodischen Verlauf der ersten Takte des Schumann-Liedes beschreiben zu können.

Nach den Ergebnissen von Krumhansl und Kessler (1982) existiert ein tonales Profil, das für jeden der zwölf Töne der chromatischen Tonleiter die Zugehörigkeitsstärke zu einer Tonart wiedergibt (vgl. a. Krumhansl, 1990). Dieses Profil entspricht in etwa den Voraussetzungen der konventionellen Musiktheorie, wonach akkordeigene Töne und dann Töne der diatonischen Tonleiter eine stärkere Zugehörigkeit zu einer Tonika haben als die übrigen Töne der chromatischen Tonleiter. Dieses tonale Profil sollte mit den Erwartungswerten der Versuchspersonen für die einzelnen Melodietöne korrelieren.

Für die melodische Kontur nimmt Schmuckler an, dass die Erwartungen für die Fortführung der melodischen Richtung stärker sein müssten als für die entgegengesetzte Bewegung. Er stützt sich hier allerdings nicht auf einzelne Ergebnisse der umfangreichen und in Kapitel 4 beschriebenen Forschungen zur melodischen Kontur.

Schmuckler findet in seinen Versuchen (1989) einen großen Einfluß der melodischen Prozesse (gap-fill und lineares Muster) und der tonalen Zugehörigkeit für die melodischen Erwartungen seiner Vpn. Die innerhalb der melodischen Prozesse folgerichtigen Töne erhielten von den Vpn. hohe Erwartungswerte. Ebenso generierten die akkordeigenen Melo-

dietöne starke Erwartungen. Die Kontur hat dagegen nach Schmucklers Auslegung keinen großen Einfluß auf die Ausbildung der Erwartungen über den Fortgang eines melodischen Verlaufs.

Die melodischen und harmonischen Erwartungen, die das Schumann-Lied als ganzes (Melodiestimme + Akkordbegleitung) hervorruft, sind nach Schmucklers Ergebnissen weitgehend voneinander unabhängig. D.h. die Erwartungseffekte, die die melodischen und harmonischen Strukturen eines Stückes generieren, sind additiv, wobei die melodischen Erwartungen generell stärker sind, wie Schmuckler innerhalb eines Regressionsmodells ermitteln konnte. Dass dennoch manchmal eine Beziehung zwischen melodischen und harmonischen Erwartungen zustande kommt, dafür sprechen nach Schmucklers Interpretation die signifikanten Korrelationen zwischen den beiden Erwartungsprofilen an drei der zehn untersuchten Stellen des Schumann-Liedes.

Die Erwartungsprofile der Versuchspersonen für Melodietöne und Akkorde benutzt Schmuckler in einem abschließenden Versuch (1989, Exp. 4), um sie mit den Improvisationen von sechs erfahrenen Pianisten in Beziehung zu setzen. In Anlehnung an die von Unyk und Carlsen (1987) verwendete Methode lässt Schmuckler die Pianisten das Schumann-Lied bis zu verschiedenen bestimmten Punkten hören, um sie direkt im Anschluß an diesen Punkt eine Fortführung improvisieren zu lassen. Schmuckler findet nicht nur eine hohe Korrelation zwischen dem ersten von den Pianisten gespielten Ton bzw. Akkord und dem melodischen bzw. harmonischen Erwartungsprofil seiner Versuchspersonen, sondern kann auch einen starken negativen Zusammenhang zwischen der Erwartungsstärke an einem bestimmten Punkt aus den Hörversuchen und der Länge der pianistischen Improvisation feststellen: Je stärker ein bestimmter Fortgang von den Hörern erwartet wurde, desto kürzer waren die Improvisationen der Pianisten. Die später publizierte inhaltliche Analyse der Improvisationen (Schmuckler, 1990) bringt leider wenig Erhellendes über die Verarbeitung der melodischen und harmonischen Details durch die Pianisten zu Tage, da Schmuckler – wie er selber bedauert – die Ebene der melodischen Motive und die Intentionen der Pianisten bei ihren Improvisationen nicht betrachtet.

Insgesamt lässt jedoch die Analyse der pianistischen Improvisationen auf einem oberflächlichen Niveau eine Konvergenz mit den Ergebnissen aus den Erwartungsbewertungen der Hörtests erkennen, die Schmuckler als weiteren Nachweis der psychischen Realität des Konzeptes der musikalischen Erwartungen interpretiert.

Einfluss von Erwartungen auf die Gedächtnisleistung:

Intuitiv liegt die Vermutung nahe, dass die Erwartungen über den musikalischen Fortgang einer Passage, die ein Hörer während des Hörens entwirft, auch für die spätere Erinnerung an diese Passage relevant sind. Denn das Aufstellen von Erwartungen ist Teil einer aktiven Verarbeitung des Gehörten, wie sie beispielsweise im *Levels-of-processing*-Ansatz (Craik & Lockhart, 1972) als relevant für die Gedächtnisleistung postuliert wird.

Wie Schmuckler (1997) allerdings bemerkt, ist die Richtung des Einflusses der Erwartungen nicht im vorhinein klar. Stark erwartete Ereignisse könnten besser erinnert werden, weil sie ein gelerntes Schema oder eine Häufigkeitsbeziehung erfüllen. Dies würde mit Erkenntnissen aus der Gedächtnisforschung zu verbalen Inhalten einhergehen, die eine erhöhte Gedächtnisleistung für prototypische Informationen bestätigen (vgl. z.B. Bartletts Untersuchungen zum Gedächtnis für Geschichten verschiedener narrativer Schemata, 1932). Allerdings ist umgekehrt auch vorstellbar, dass gerade seltene und ungewöhnliche Ereignisse besser in der Erinnerung bleiben, da deren unerwartetes Eintreten eine erhöhte Aufmerksamkeit und eine tiefere Verarbeitung zur Folge haben könnten. Letztere Alternative bezieht Schmuckler auf den bekannten von-Restorff-Effekt (von Restorff, 1933; Zusammenfassung bei Wallace, 1965).

Um den Zusammenhang zwischen Erwartung und Gedächtnisleistung zu prüfen, erhebt Schmuckler in seiner Versuchsreihe sowohl die subjektive Erwartung der Schlusstöne von einstimmigen Melodien als auch die Gedächtnisleistung für diese Schlusstöne in einem späteren Wiedererkennenstest. Er findet empirische Belege nur für den positiven Zusammenhang zwischen der starken Erwartung und dem besseren Gedächtnis für melodische Ereignisse. Der von-Restorff-Effekt lässt sich nach den Ergebnissen von 1997 nicht belegen. Dabei muss Schmuckler allerdings eingestehen, dass durch die Wahl der Varianten in den Melodie-Schlusstönen keine extrem ungewöhnlichen melodischen Ereignisse in seinem Stimulusmaterial vorhanden waren. Insofern ist eine U-förmige Beziehung zwischen Erwartetheit und Gedächtnisleistung nicht auszuschließen, nach der sehr unerwartete (überraschende) und sehr erwartete melodische Ereignisse besonders gut erinnert werden.

Das zweite zentrale Erkenntnisinteresse von Schmucklers Studie zur Beziehung zwischen Erwartungen und Gedächtnis (1997) betrifft die Symmetrie der Verwechslungen von Melodien im Gedächtnis. Dies betrifft die Frage, ob stark erwartete Melodievarianten genauso häufig mit wenig erwarteten verwechselt wie umgekehrt?

Die empirischen Ergebnisse zeigen eine deutliche Asymmetrie bei den Verwechslungen im Wiedererkennenstest: Ist die später gehörte Testmelodie stärker erwartet als die davor ge-

hörte Referenzmelodie, so zeigten die Vpn. eine starke Tendenz, die Testmelodie als identisch zur Standardmelodie zu bewerten, auch wenn sich die beiden Melodien unterschieden (Schmuckler, 1997, Exp. 2). Anders formuliert: Melodien, die näher an einem erwarteten Schema liegen, rufen eher das Gefühl einer Erinnerung hervor als weniger erwartete Melodien und zwar bis zu einem gewissen Grade unabhängig davon, ob sie tatsächlich direkt zuvor gehört wurden. Im Gedächtnis vermischen sich also Erwartetes/Schematisches und Bekanntes.

Neben der subjektiven Erwartetheit bestimmt Schmuckler mit einem von ihm und Krumhansl entworfenen Algorithmus zusätzlich die von den Melodieenden implizierte Tonalität. Diese Tonalität am Melodieende setzt er mit der Tonalität der jeweiligen Melodieanfänge in Beziehung. Die Entfernung zwischen den Tonalitätseindrücken von Melodieanfang und -ende verwendet er ebenso wie die Variable *Stärke der Erwartung* (expectedness) als Prädiktorvariable, um die Gedächtnisleistungen seiner Versuchspersonen zu erklären. Er findet jedoch nur in einem der beiden Experimente (Exp. 1) einen signifikanten Zusammenhang zwischen der ermittelten Tonalitätsdistanz und den Gedächtnisleistungen. Und selbst zusammen mit der Erwartetheit klärt die tonale Distanz nur 48% der Varianz in einem multiplen Regressionsmodell auf. Schmuckler vermutet, dass ein weiterer signifikanter Einflussfaktor bei der Erklärung der Gedächtnisleistungen die melodische Kontur, oder – besser gesagt – die Ähnlichkeit zwischen den Konturen von Referenz- und Testmelodien sein könnte. Jedoch nimmt er kein Maß für die Konturähnlichkeit in das Regressionsmodell dieser Studie auf.

Auch in seiner gemeinsamen Studie mit Penny Chiappe (Chiappe & Schmuckler, 1997) betrachtet Schmuckler Gedächtnisleistungen. Jedoch verwendet er nicht die Erwartetheit als Prädiktorvariable, sondern geht von der Phrasensegmentierung der Testmelodien aus. Dennoch sollen die Ergebnisse dieser Studie hier kurz referiert werden, da sie für den Bereich des Musikgedächtnisses interessant sind.

Die Einteilung einer Melodie in Phrasen kommt nach Chiappe und Schmuckler hauptsächlich durch melodische Sprünge und deutliche Unterschiede in den Tondauern zustande. Sie stützen sich dabei auf empirische Erkenntnisse von Krumhansl und Jusczyk (1990; Jusczyk & Krumhansl, 1993), die an den Stellen, an denen Hörer Phrasenenden festlegten, eine Zu- und dann Abnahme der Tondauern um das Phrasenende herum sowie einen starken Abfall mit anschließendem Anstieg in der Tonhöhe feststellen konnten. Leider geben Chiappe und Schmuckler jedoch keinen Algorithmus für die Ermittlung von Phrasenenden an, sondern

nehmen die Einteilung ihrer Experimentalmelodien in Phrasen intuitiv nach den Erkenntnissen von Krumhansl und Juszcyk vor (Chiappe & Schumckler, 1997, S. 255).

Neben dem erwarteten Ergebnis, dass die Phraseneinteilung einer Melodie generell das Gedächtnis für Melodien beeinflussen kann, erscheinen zwei weitere – eigentlich nebensächliche – Ergebnisse gleichermaßen interessant:

Erstens konnten Chiappe und Schmuckler den Einfluss der Phrasierung auf das Gedächtnis nur bei ihren musikalisch gut vorgebildeten Versuchspersonen beobachten. Die Versuchspersonen, die kein musikalisches Training genossen hatten, schienen die Phraseneinteilungen für den Gedächtnistest nicht zu verarbeiten. Dies lässt sich in der mittlerweile langen Reihe von empirischen Hinweisen lesen (Sloboda & Parker, 1985, und die oben angeführten Nachfolgestudien; Dowling et al, 1995;), die zeigen, dass musikalisch nicht aktive Versuchspersonen die musikalischen Strukturen der Versuchsstimuli in der Regel nicht zur Verbesserung der eigenen Gedächtnisleistung verwenden können.

Zweitens zeigte sich ein deutlicher *recency*-Effekt, nach dem Veränderungen in den letzten drei bzw. vier Tönen deutlich besser erkannt wurden als in den jeweils vorletzten drei bis vier Tönen. Ein solch deutlicher *recency*-Effekt, wie er beispielsweise aus dem verbalen Lernen bekannt ist, ist zuvor mit echten Melodien noch nicht beobachtet worden und war für die nicht musikalisch vorgebildeten Versuchspersonen in diesem Experiment sogar noch deutlicher ausgeprägt. Dieser *recency*-Effekt mag auf das extrem kurze Retentionsintervall von 1 Sekunde zurückgehen, welches die Referenzmelodie von der Testmelodie im Wiedererkennenstest trennte.

Zusammenfassung und Diskussion:

Im Gegensatz zu Modellen von Monahan und Boltz und Jones stellt das Erwartungskonzept von Schmuckler nur einen mittelbaren Zusammenhang zwischen Struktureigenschaften des musikalischen Materials und Gedächtnisleistungen her. Mittler zwischen jenen Einflüssen und diesen Effekten sind die musikalischen Erwartungen.

Erwartungen werden nach Schmucklers theoretischem Entwurf und seinen empirischen Ergebnissen durch die Häufigkeit von Akkordverbindungen, durch prototypische melodische Prozesse im Sinne von Leonard Meyer, durch harmonische Beziehungen der Melodietöne und wahrscheinlich in geringerem Maße durch den Verlauf der melodischen Kontur generiert. Auf der anderen Seite findet Schmuckler einen starken positiven Zusammenhang zwischen den Erwartungen für ein melodisches Ereignis und den Behaltensleistungen

für dieses Ereignis, von dem er zumindest implizit annimmt, dass er linear sei. Die Stärke der Hörerwartung kann somit zur Vorhersage von Gedächtnisleistungen dienen.

Es gibt jedoch eine Reihe von Punkten und ungeklärten Fragen, die eine Verwendung des Erwartungskonzeptes für die Vorhersage und Analyse von Gedächtnisleistungen in größeren Zusammenhängen sehr erschweren oder vielleicht sogar unmöglich machen:

Schmuckler selbst gesteht ein, dass die von ihm verwendete Häufigkeitstabelle der Akkordverbindungen von Piston seine Ergebnisse nur teilweise erklären kann. Die Tabelle müsste inhaltlich korrigiert und auf größere harmonische Zusammenhänge (Markov-Ketten höherer Ordnung) ausgeweitet werden. Eine solche Tabelle dürfte nur sehr aufwendig herzustellen sein und existiert weder nach Schmucklers noch nach dem Wissen des Autors.

Der Rhythmus bzw. die zeitliche Dimension von musikalischen Ereignissen ist nicht integraler Bestandteil des Erwartungskonzeptes, obwohl Schmuckler und Boltz (1994) die zeitliche Regelmäßigkeit von Akkordeinsätzen als einflussnehmenden Faktor feststellen.

Der Verlauf einer melodischen Kontur kann nach Schmuckler einzig die Erwartung einer Konturfortführung generieren. Dies scheint jedoch schon in Kenntnis von bekannten Stimmführungsregeln, nach denen beispielsweise auf einen Sprung eine schrittweise Gegenbewegung folgen sollte (vgl. z.B. de la Motte, 1976, S. 22), zu einfach gedacht. Auch nach den Modellen von Monahan sowie Boltz und Jones sind differenziertere Konturverläufe erwartbar, bei denen die Erwartung etwa durch rhythmische Strukturen generiert werden.

Theoretisch ist die Rolle der Tonalitätszugehörigkeit von Melodietönen in Schmucklers Konzept nicht geklärt. Während sie in der großen Studie von 1989 als ein Faktor betrachtet wird, der zur Generierung von Erwartungen beiträgt, verwendet Schmuckler 1997 die Tonalitätszugehörigkeit als zusätzlichen Faktor, der parallel mit den Erwartungswerten in das Regressionsmodell zur Erklärung der Gedächtnisleistungen eingeht.

Genauso unklar ist, wie die Zugehörigkeit eines Melodietones zu einem melodischen Schema oder Prototyp automatisch bestimmt werden soll. Diesen Analyseschritt vollzieht Schmuckler (1989) als Mensch mit geeignetem Musikverständnis. Ein Formalismus, der die melodischen Prozesse in einer Melodie entdeckt und einzelnen Tönen eine Funktion innerhalb eines Schemas zuweist, scheint jedoch ebenso wenig zu existieren wie ein vollständig automatisierbarer Algorithmus für die Tonalitätsbestimmung oder für die harmonische Analyse in Schmucklers Modell. Die Bestimmung der melodischen und harmoni-

schen Erwartungswerte für eine große Menge von Melodien ist deshalb extrem zeitaufwendig und nicht praktikabel.

Schließlich ist das Konzept der musikalischen Erwartungen, wie es Schmuckler darstellt, als Ganzes nicht unproblematisch. Schmuckler geht von der (nur implizit geäußerten) Voraussetzung aus, dass die aufgezählten Faktoren bei allen Versuchspersonen die gleichen Erwartungen für bestimmte Passagen oder musikalische Ereignisse erzeugen. Wie Unyk und Carlsen in ihrer viel zitierten Studie (1987) jedoch feststellen mussten, können melodische Erwartungen individuell sehr verschieden bei den einzelnen Versuchspersonen ausgeprägt sein: „if [...] the content of these expectancies are found to differ across subjects, then a measure of the effect of violation of expectancy must also reflect these individual differences.“ (Unyk & Carlsen, 1987, S. 14). Konsequenterweise gehen Unyk und Carlsen in ihren Analysen der melodischen Erwartungen nur von individuellen Erwartungsprofilen aus. Möglicherweise gehen die individuellen Unterschiede in den Erwartungsbewertungen auch auf ein unterschiedliches Verständnis des Konzeptes musikalischer Erwartungen bei den Vpn. zurück. Immerhin ist der Begriff *musikalische Erwartung* kein Terminus der alltäglichen Sprache, mit der Musik beschrieben wird, und auch als musiktheoretischer Terminus ist ‚Erwartung‘ eher ungebräuchlich. Da die Erwartungen von Schmuckler ausnahmslos über subjektive *ratings* erhoben werden, ist nicht auszuschließen, dass die einzelnen Versuchsteilnehmer ihre Bewertungen nach unterschiedlichen subjektiven Konzepten vornahmen. Eine methodologische Alternative stellt beispielsweise die indirekte Bestimmung von Erwartungen über perzeptuelle Restaurationsleistungen für teilweise gestörte musikalische Strukturen dar, die DeWitt und Samuel (1990) mit einer eleganten Versuchsreihe demonstriert haben. Da es sich bei den von Schmuckler (1989, 1990, 1997; Schmuckler & Boltz, 1994) untersuchten Versuchspersonen ausschließlich um musikalisch gut und sehr gut vorgebildete Personen handelte, dürften die subjektiven Erwartungskonzepte der Versuchspersonen zumindest noch ähnlich gewesen sein. Eine Bestätigung des Erwartungskonzeptes mit musikalisch nicht vorgebildeten Versuchspersonen steht noch aus, und es sind bei dieser Gruppe von Personen noch größere Unterschiede in den subjektiven Erwartungskonzepten und dadurch geringere Effekte der Erwartung generierenden Faktoren zu erwarten. Einen Zusammenhang zwischen der musikalischen Vorbildung und den Leistungen in erwartungsgesteuerten Unterscheidungstests konnten z.B. Dewitt und Samuel schon 1990 nachweisen (S.141).

Die vorangegangenen Kritikpunkte zeigen, wie schwierig der Einsatz des Erwartungskonzeptes für die Vorhersage von Gedächtnisleistungen aufgrund von Struktureigenschaften eines Musikstücks oder einer Melodie sein dürfte. Die Repräsentation einer Melodie nach den von Schmuckler vorgeschlagenen Analysestrukturen und -kriterien scheint nicht ohne weiteres vollständig formalisierbar zu sein. Aus dieser Repräsentation sind Gedächtnisleistungen nur über den ‚Umweg‘ des Erwartungskonzeptes vorhersagbar. Und schließlich ist sowohl die Quantifizierung der erwartungsgenerierenden Faktoren sowie der Erwartungsstärke und der davon abhängigen Gedächtnisleistungen (bisher) nicht adäquat möglich, wie Schmuckler selbst eingesteht (1997). Zieht man zu guter Letzt noch in Betracht, dass nur eine einzige Studie bislang den Zusammenhang von Erwartungen und Gedächtnisleistungen nach Schmucklers Konzept belegt hat, so erscheint eine eingehendere Beschäftigung mit Schmucklers Konzept der musikalischen Erwartungen mit der Perspektive der Vorhersage von Gedächtnisleistungen vorerst nicht sinnvoll. Für die Erklärung anderer kognitiver Mechanismen, wie beispielsweise der Restauration von musikalischen Informationen in teilweise gestörten Musiksignalen, kann das Konzept der musikalischen Erwartung in der Formulierung von Schmuckler oder anderen Autoren (z.B. Carlsen, 1981; Dewitt & Samuel, 1990) sicher erkenntnisgewinnend eingesetzt werden.

3.3.4 Melodieexperimente zur Überprüfung von allgemeineren Gedächtnistheorien

Die in diesem Unterabschnitt vorgestellten Studien haben weder einen gemeinsamen theoretischen Forschungsansatz noch bedienen sie sich der gleichen experimentellen Methodik. Es verbindet sie lediglich die Verwendung von kurzen Melodien oder von liedhaften melodischen Sequenzen als Versuchsstimuli und die Tatsache, dass im Ergebnis Aussagen über das melodische Gedächtnis gemacht werden, die im Zusammenhang dieser Arbeit interessant erscheinen. Es soll hier deshalb für jede Studie eine ergebnisorientierte Zusammenfassung gegeben werden, die kurz den Hintergrund der Studie und das für das Melodiegedächtnis interessante Resultat wiedergibt. In Stichworten sind die einzelnen Ergebnisse zusätzlich in der tabellarischen Zusammenfassung des folgenden Teilkapitels aufgeführt.

3.3.4.1 Dissoziation zwischen semantischem und episodischem Gedächtnis für Melodien

Die Studie von Gardiner und Kollegen (Gardiner et al., 1996) geht von dem Modell des episodischen Gedächtnisses nach Tulving (z.B. Tulving, 1972, 1983) aus, wonach das

Langzeitgedächtnis (LZG) neben semantischem und prozeduralem Gedächtnis als dritte Komponente das episodische Gedächtnis enthält, in dem räumlich und zeitlich datierbare Ereignisse und persönliche Erfahrungen gespeichert sind (vgl. Vaterrodt-Plünnecke, 2001, S. 142). Die Arten der Erinnerung an semantische und episodische Gedächtnisinhalte lassen sich gut unterscheiden und werden von Tulving als „know“ bzw. „remember responses“ bezeichnet (Tulving, 1985). Im Kontext von vielen vorausgegangenen Studien belegen Gardiner und Mitarbeiter nun mit Hilfe eines Wiedererkennensexperimentes mit Ausschnitten aus echten Melodien (polnische Volkslieder und Themen aus klassischen Werken), dass die beiden Arten der Erinnerung – und damit implizit auch die semantische und episodische Gedächtniskomponente – vollständig unabhängig voneinander sind. Während alle übrigen Muster der Entwicklung von *remember* und *know responses* in Abhängigkeit von einem gemeinsamen Einflussfaktor schon zuvor belegt worden waren, zeigen Garinder et al. in den Experimenten 1 und 3, dass unter gewissen Umständen *remember* und *know responses* durch den Einfluss eines gemeinsamen Faktors auch beide steigen können: Beim Wiedererkennen der polnischen Volkslieder stiegen die *remember* und *know responses* mit der Anzahl von Durchgängen, in denen die Versuchspersonen die Volkslieder gehört hatten. Bei den Themen der klassischen Musik dagegen stiegen nur die *remember*-Angaben mit der Zahl der Wiederholungen, während die *know*-Angaben stagnierten. Als mögliche Erklärung geben die Autoren an, dass die Speicherung im semantischen Gedächtnis generell von einem etablierten Schema für eine Klasse von Objekten abhängt. Möglicherweise war ein Schema für die gängigeren klassischen Melodien bei den Versuchspersonen schon vor Beginn des Versuchs ausgeprägt, während sich ein Schema für die Verarbeitung und Speicherung der weniger geläufigeren polnischen Volksmelodien erst im Laufe der bis zu vier Wiederholungen angelegt bzw. entsprechend angepasst wird.

Eine Parallele lässt sich von diesem Ergebnis zu den Experimenten von Oura und Hatano (1988) sowie Zielinska und Miklaszewski (1992) ziehen (s.o.), die die Bekanntheit mit dem Stil einer Melodie bzw. mit dem melodischen Strukturschema als stark einflussnehmende Variable für die Wiedergabe von unbekannten Melodien identifiziert hatten. So scheint auch diese Studie, die eigentlich der Überprüfung der Modells des episodischen Gedächtnisses diente, die Bedeutung von kognitiven Schemata für die Melodieverarbeitung und -erinnerung zu unterstreichen.

3.3.4.2 Untersuchungen zum Gedächtnis im Alter anhand von Melodien

Auch in der Studie von Bartlett, Halpern und Dowling (Bartlett et al., 1995) spielt das episodische Gedächtnismodell nach Tulving eine Rolle bei der Erklärung der Versuchsergebnisse. Ziel dieser Studie war jedoch eigentlich die Charakterisierung von Gedächtnisleistungen bei Alzheimer Patienten. Wie Bäckman und Herlitz (1990) schon ausführlich experimentell dokumentiert hatten, können Alzheimer Patienten nicht wie Gesunde von bekannten Tatsachen in Wiedererkennensexperimenten profitieren. Der sogenannte „Mirror-Effekt“ – die zunehmende richtige Identifizierung von vorbekannten Objekten im Wiedererkennenstest und die gleichzeitige geringere falsche Identifizierung von diesen Objekten (vgl. Hintzman, 1994) – blieb bei Alzheimer-Patienten aus. Bartlett und Kollegen untersuchten nun das Auftreten des *Mirror*-Effektes bei Alzheimer-Patienten und gesunden älteren Menschen (Altersdurchschnitt 70 Jahre) in einem Wiedererkennenstest mit bekannten und unbekannten Melodien und verglichen die Gedächtnisleistungen mit denen von jüngeren gesunden Versuchspersonen (Studenten). Die älteren wie die jüngeren Gesunden zeigten den erwarteten *Mirror*-Effekt, d.h. bei den bekannten Melodien gaben sie häufiger richtigerweise an, dass sie die Melodien direkt zuvor in der Testsession gehört hatten und behaupteten seltener fälschlicherweise, dass sie eine bekannte Melodie direkt zuvor gehört hatten, wenn dies nicht der Fall war. Die Wiedererkennensleistungen für bekannte Melodien waren bei Gesunden also auf beiden Maßen (*hits* und *false alarms*) besser. Generell erreichten die jüngeren Versuchspersonen bessere Wiedererkennensleistungen. Die Alzheimer-Patienten zeigten dagegen bei den bekannten Melodien eine eindeutige Tendenz anzugeben, diese direkt zuvor gehört zu haben, unabhängig davon, ob sie sie wirklich zuvor gehört hatten oder nicht. Verglichen mit den Leistungen der gesunden Versuchspersonen zeigten die Patienten größere Defizite bei den bekannten Melodien als bei den unbekannten.

Bartlett und Kollegen interpretieren das Ausbleiben des *Mirror*-Effektes bei den Patienten dahingehend, dass die Gedächtnisdefizite bei an Alzheimer Erkrankten qualitativ von denen älterer Versuchspersonen verschieden seien. Es existiere also kein Kontinuum von den Gedächtnisleistungen junger Versuchspersonen über normale Altersdefizite zu den Leistungsausfällen bei Alzheimerkranken. Bartlett, Halpern und Dowling konnten darüber hinaus bei jüngeren und älteren Versuchspersonen zwei verschiedenen Gedächtnisstrategien für den Umgang mit unbekannten und bekannten Melodien feststellen. Während die Älteren zu den bekannten Melodien i.d.R. den Titel erinnerten und die Melodie so verbal etikettieren konnten (Rekodierung im semantischen Gedächtnis), speicherten die Jüngeren

dagegen eher die Erinnerung an die Lernepisode in der Lernsession selbst, und verwendeten so ihr episodisches Gedächtnis. Die Alzheimer-Patienten konnten offenbar jedoch weder die eine noch die andere Strategie anwenden. Dies spricht nach den Autoren der Studie dafür, dass Alzheimerkranke ihr eigenes Wissen nicht nutzen können. Denn ein Wissen um die Bekanntheit der allgemein bekannten Melodien scheint auch bei ihnen vorhanden zu sein, da sich nur so die Tendenz bei den Erkrankten erklären lässt, eine bekannte Melodie als direkt zuvor gehört zu identifizieren.

Im Zusammenhang mit der vorliegenden Arbeit lassen sich aus dieser Studie zwei Erkenntnisse festhalten: Erstens sind die Leistungen für das Wiedererkennen von Melodien bei jüngeren Versuchspersonen besser als bei älteren. Dies mag mit der möglicherweise effektiveren episodischen Gedächtnisstrategie zu tun haben, die die jüngeren Versuchspersonen benutzten. Zweitens werden insgesamt beim Wiedererkennen von vorher bekannten Melodien weniger Fehler bzw. mehr richtige Identifizierungen gemacht als beim Wiedererkennen von unbekannten Melodien.

Eine Folgestudie von Halpern, Bartlett und Dowling (Halpern et al., 1995) überprüfte die gleichzeitigen Einflüsse von Alter und musikalischer Expertise in einem Wiedererkennungsexperiment mit transponierten Melodien. Auch in diesem Experiment zeigten die älteren Versuchsteilnehmer insgesamt schlechtere Leistungen beim Wiedererkennen der Melodien. Die Autoren fanden zusätzlich nur in einem der vier berichteten Experimente eine erwartete Interaktion dergestalt, dass eine größere musikalische Vorbildung die negativen Effekte des Lebensalters auf die Gedächtnisleistung abschwächen oder sogar aufheben konnte. In anderen Bereichen, z.B. den Anschlagsgeschwindigkeiten von Schreibkräften (Salthouse, 1984) oder dem Gedächtnis von Piloten für Luftverkehrsmeldungen (Morrow et al., 1992, 1994), waren die entgegengesetzten Effekte von Alter und Erfahrung jedoch bereits öfter nachgewiesen worden.

Die Interpretation, die Mainz und Salthouse (Mainz & Salthouse, 1998) zu den Ergebnissen von Halpern und Kollegen geben, geht deshalb davon aus, dass die von Halpern verwendete Wiedererkennungsaufgabe nur in sehr geringem Maße das Vorwissen der getesteten Musiker forderte. Aus diesem Grund testeten Mainz und Salthouse Personen aller Altersklassen mit einer Gedächtnisaufgabe, die in der Wiedergabe von kurzen, visuell präsentierten Melodien (eigentlich: kurzen Tonfolgen) in einem Notensystem bestand und das musikalische Vorwissen sowie die Erfahrung mit musikalischer Notation stärker beanspruchte. Erwartungsgemäß konnten die musikalisch aktiveren und die mit der Notation

besser vertraute Versuchspersonen die Aufgabe besser bewältigen. Auch ein negativer Effekt des Lebensalters zeigte sich bei der Analyse der Versuchsergebnisse. Zur Enttäuschung der beiden Autoren gaben die Daten keinen Anhaltspunkt für eine signifikante Interaktion zwischen Alter und musikalischer Erfahrung: Auch langjährige musikalische Erfahrung und aktives Musizieren konnte den negativen Effekt des Alters auf das Melodiegedächtnis nicht aufhalten. Mainz und Salthouse vermuten zwar, dass das Computerinterface, welches sie zur Durchführung des Tests verwendeten, zum Teil für die schlechteren Leistungen der älteren Versuchsteilnehmer verantwortlich gewesen sein könnte, da diese generell geringere Erfahrung im Umgang mit dem Computer hätten. Dennoch bleibt zu konstatieren, dass für den Bereich der Musik ein Ausgleich der altersbedingten Gedächtnisdefizite durch musikalische Erfahrung bisher nicht nachgewiesen werden konnte.

3.3.5 Empirische Ergebnisse der Forschungen zum Gedächtnis für Melodien

(Tabelle)

Ergebnis	Kommentar / Interpretation	Beleg
Wiedergabe unbekannter Melodien (Kap. 3.2.1)		
Bei mehrfacher Präsentation und Wiedergabe (Singen) steigen Anzahl wiedergegebener Töne und falscher Töne im gleichen Maß bis zu einem Wendepunkt an; nach Wendepunkt nimmt Fehleranteil an Wiedergaben ab.		Sloboda & Parker (1985); Zielinska & Miklaszewski (1992)
Wendepunkt wird durch Schwierigkeit der Melodie und Erfahrung der Vpn. mit Stil von Melodie bestimmt.	Je einfacher und kürzer die Melodie und je erfahrener die Vpn. ist, desto eher wird Wendepunkt erreicht bzw. Melodie vollständig gelernt.	Oura & Hatano (1988); Zielinska und Miklaszewski (1992); Ogawa et al. (1995); Gardiner et al. (1996)
Wenn Vpn. mit Stil von Melodie vertraut sind und gewisse musikalische Expertise haben, machen sie i.d.R. ‚musikalisch sinnvolle‘ Fehler.	Interpretation (Sloboda & Parker; Zielinska & Miklaszewski): Struktur von Melodie wird zuerst gelernt, Details z.T. improvisiert.	Sloboda & Parker (1985); Oura & Hatano (1988); Zielinska und Miklaszewski (1992)
Die Skelettstruktur von Melodien wird zumindest von musikalisch vorgebildeten Personen vorrangig behalten.		Large et al. (1995)
Vpn. ohne musikalische Expertise zeigen andere, schwer systematisierbare Fehler: Konturfehler, Modulatio-		Oura & Hatano (1988)

nen, fehlende Segmentierung, Fixierung auf Endphrase.		
Wiedererkennen bekannter manipulierter Melodien (Kap. 3.2.2.1)		
Wenige Töne (6-12) reichen aus, um bekannte Melodie auch mit z.T. starken Veränderungen zu erkennen.		White (1960); Idson & Massaro (1978)
Melodische Kontur und relative Intervallgröße sind für Wiedererkennen von großer Bedeutung.		White (1960); Dowling & Fujitani (1971, Exp. 2); Idson & Massaro (1978)
Tonchroma wird für Wiedererkennen verwendet, wenn melodische Kontur intakt ist.		Dowling & Hollombe (1977); Idson & Massaro (1978)
Tonhöhenverlauf ist stärkerer Gedächtnis-cue für Wiedererkennen als Rhythmus.		White (1960); Hébert & Peretz (1997)
Verbale Vorinformation kann Wiedererkennen stark unterstützen.	Mit verbaler Vorinformation wird offenbar andere Strategie (<i>Top-Down-Prozess</i> aktive Suche) benutzt;	Deutsch (1972); Dowling (1973b, Exp. 3); Hébert & Peretz (1997)
Melodiestructur und Subjektfaktoren spielen in Wiedererkennensprozess bedeutende Rolle.		White (1960); Dowling & Hollombe (1977); Idson & Massaro (1978)
Speicherung von unbekannten Melodien (3.2.2.2)		
Varianten in unbekannten Melodien werden schlechter erkannt als in bekannten, aber deutlich besser als in künstlichen Tonfolgen.	Echte Melodien sind (u.a. durch Rhythmus) reichere Reizmuster als kurze Tonfolgen (Dowlings Interpretation).	Dowling et al. (1995);
Bei unbekannten Melodien verlassen sich Personen ohne musikalische Vorbildung stärker auf Gedächtnis für absolute Tonhöhen.	Die musikalische Vorbildung beeinflusst die Speicherung der Struktur von unbekannten Melodien positiv (eigene Interpretation).	Dowling et al. (1995)
Eröffnungsphrase eines Stückes wird am besten erinnert.	Eröffnungsphrase hat Wahrnehmungspriorität.	Dowling et al. (2002)
Wenn Stille statt Musik auf zu speichernde Phrase folgt, wird diese besser erinnert.	Letzte Phrase eines Stückes wird gut erinnert.	Dowling et al. (2002)
Die Repräsentation von melodischen Phrasen wird im fortlaufenden Hörprozess präziser.	Automatische Verarbeitung von gehörten Melodien beim Hören, implizites Gedächtnis (Interp: Dowling); kontinuierlicher Wandel der Gedächtnisrepräsentation im Hörprozeß.	Dowling et al. (2002)
Vorwissen um Gedächtnistest erhöht	Willentliche Aufmerksam-	Dowling et al. (2002)

Gedächtnisleistung erheblich.	keitsfokussierung für Gedächtnis von Bedeutung, explizites Gedächtnis.	
Erinnerungsleistung an melodische Passagen nimmt mit Zeit entlang negativer Exponentialfunktion ab.	Vergessenskurve ähnlich der für verbales Material (s. z.B. Anderson, 1995).	Kauffman & Carlsen (1989)
Vergessen von Musikexperten und –laien folgt ähnlicher Funktion, aber Experten zeigen absolut bessere Leistungen.		Kauffman & Carlsen (1989)
Erinnerung an Ausschnitt echter Musik von 7-8 Jährigen bleibt für Wiedererkennungsaufgabe über 40 Sek. stabil, wenn Stille in R-Intervall.		Kauffman (1990)
Bekanntheit mit musikalischem Stil verbessert Erinnerungsleistungen für Stücke dieses Stils.	Vgl. Oura & Hatano (1988), Kap. 3.2.1.	Russell (1987)
Kein Zusammenhang zwischen subjektiv geschätzter Erinnerbarkeit und tatsächlichen Gedächtnisleistungen für ein Stück.	Gilt zumindest für Stücke aus musikalischem Stil, mit dem Vpn. nicht besonders vertraut sind.	Russell (1987)
Subjektiv geschätzte Komplexität korreliert positiv mit Tempo des Stückes.		Russell (1987)
Verbinden von konkreten Vorstellungsbildern mit gehörten Musikausschnitten verbessert Erinnerungsleistung für Ausschnitte.		Delis et al. (1978)
Bedeutung von regelhaftem Aufbau und Erwartung (Kap. 3.2.3)		
Töne, bei denen Akzente aus rhythmischer, (harmonischer) und Tonhöhen-dimension zusammenfallen, werden besser enkodiert und erinnert.	Ermittlung von Akzenten in den einzelnen Dimensionen sowie Voraussagen zur Erinnerbarkeit sind nach den ähnlichen Modellen von Boltz und Jones, Monahan oder Drake et al. möglich.	Boltz & Jones (1986); Jones (1987); Monahan et al., (1987); Drake et al. (1991); Boltz (1993)
Die relative Verteilung von melodischen und rhythmischen Akzenten bestimmt die kognitive Ähnlichkeit von Melodien.	Theorie von Boltz und Jones; ist für Melodien im Kontext ‚echter‘ Musik noch nachzuweisen.	Boltz & Jones (1986); Jones (1993)
Musikalische Vorbildung beeinflusst Verarbeitung und Speicherung von Strukturen in rhythmischer und Tonhöhendimension positiv.		Monahan et al., (1987); Smith & Cuddy (1989)
Subjektive Erwartungen für ein musikalisches Ereignis (z.B. Fortgang einer Melodie) und Gedächtnisleistungen für dieses Ereignis korrelieren stark positiv miteinander.	Subjektive Erwartungen können anhand von melodischen und harmonischen strukturellen Parametern vorhergesagt werden (z.B. Schmuckler,	Unyk & Carlsen (1987); Schmuckler (1997)

	1989, 1990).	
3.2.4 Melodieexperimente zur Überprüfung von allgemeineren Gedächtnistheorien		
Die Leistungen beim Wiedererkennen von Melodien nehmen mit dem Lebensalter ab.		Bartlett et al. (1995); Meinz & Salthouse (1998)
Alzheimerpatienten können ihr eigenes Wissen um bekannte Melodien nicht verwenden und sind in Wiedererkennenstest schlechter als Gesunde.		Bartlett et al. (1995)
Wiedererkennensleistungen für vorbekannte Melodien sind deutlich höher als für unbekannte.		Bartlett et al. (1995)
Sind Vpn. mit dem Stil einer Melodie vertraut, so kann diese Melodie im episodischen, wie im semantischen Gedächtnis verarbeitet werden; sind sie nicht mit dem Stil vertraut, so wird sie v.a. episodisch gespeichert.		Bartlett et al. (1995); Gardiner et al. (1996)
Die Speicherung im episodischen Gedächtnis ist in Wiedererkennenstest u.U. effektiver.	Eigene Interpretation der Ergebnisse von Bartlett et al. (1995).	Bartlett et al. (1995)

4 Melodische Ähnlichkeit

4.1 Melodische Ähnlichkeit als Werkzeug zur Analyse von Gedächtnisleistungen und zur Modellierung von Gedächtnismechanismen

Die Ähnlichkeit von Melodien oder melodischen Passagen ist für viele Bereiche und Aufgabenstellungen der Musikwissenschaft ein zentrales Konzept. Deshalb ist es wenig verwunderlich, dass zur Bestimmung von melodischer Ähnlichkeit für die verschiedenen Domänen eine ganze Reihe von Ähnlichkeitsmodellen bzw. -funktionen entwickelt wurden, die sich im Ansatz und in der mathematischen Konstruktion zum Teil deutlich unterscheiden. Zu den Domänen, für die die Bestimmung von melodischer Ähnlichkeit zentral ist und in denen entsprechende elaborierte Konzepte erarbeitet wurden, gehören die folgenden fünf Bereiche:

- a) Musikethnologie und Volksliedforschung
- b) Musikanalyse
- c) Musikwissenschaftliche Beschäftigung mit Urheberrechtsfragen
- d) Retrieval Systeme für Melodiedatenbanken
- e) Musikpsychologie

Diese Bereiche sollen im folgenden kurz umrissen werden. Dabei ist ein etwas größeres Augenmerk auf die Bereiche c) und d) gelegt, die nicht zu den klassischen Kernfeldern musikwissenschaftlicher Beschäftigung zählen.

4.1.1 Anwendungsfelder melodischer Ähnlichkeitsmodelle

a) Musikethnologie und Volksliedforschung:

Zentrale Forschungsanliegen sind die Identifizierung von zusammengehörigen Varianten einer Melodie oder eines Liedtypes in der Volksliedforschung und die Einteilung von großen Melodiecorpora in Melodietypen. Im Idealfall bekommen die anhand von Ähnlichkeitsbestimmungen gefundenen Melodietypen eine externe Validität durch die sinnvolle Zuordnung zu geographischen Regionen, musikalischen Funktionen oder Epochen der Entstehung bzw. Komposition.

Ein Beispiel für den ersten Fall stellen die Kriterien der Variantensuche innerhalb der Verfahren zur interaktiven Melodieanalyse von Barbara Jesser dar (Jesser, 1990, S. 208ff). Ebenso lassen sich melodische Varianten gut identifizieren mit den Verfahren, die Stein-

beck für den Vergleich von „zeitabhängigen Tonhöhenverläufen“ vorschlägt (Steinbeck, 1982, S. 102-131) und die der Mathematiker Manfred Leppig aufgreift und differenziert (Leppig, 1987a, 1987b).

Die Beispiele für die Klassifikation von Volksliedsammlungen sind zahlreicher. Als bekannte Projekte lassen sich Bartóks Ordnung der Parry-Sammlung von serbokroatischen Volksliedern (Bartók & Lord, 1951; zum methodologischen Hintergrund s.a. Bartók, 1976 [ungar. Orig. 1936]), die faktorenanalytische Typenbestimmung von Volksliedmelodien von Reiner Kluge (Kluge, 1974), die methodisch ähnliche Typenbestimmung an ungarischen Volksmelodien mit Hilfe der Hauptkomponentenanalyse bei Juhász (2000), die auf Clusteranalysen beruhenden Einteilungen von Volksliedern durch Wolfram Steinbeck (1982) oder die nach unterschiedlichen Kriterien geordneten Klassifizierungen des Melodiebestandes des Deutschen Volksliedarchives (z.B. Wiora, 1977; Suppan & Stief, 1976; Jesser, 1990) nennen.

Da im Zusammenhang dieser Arbeit eine Konzentration auf Methoden zur Ähnlichkeitsbestimmung für melodische Varianten geboten ist, können die letztgenannten Ansätze zur Klassifikation großer Melodiensammlungen leider nur vereinzelt und am Rande eine Berücksichtigung erfahren.

b) Musikanalyse:

Der Vergleich von melodischen Gebilden spielt in der Musikanalyse generell eine große Rolle. Sowohl die traditionelle Musikanalyse kommt durch den Vergleich von ähnlichen melodischen Strukturen zu Erkenntnissen über Aufbau und kompositorische Anlage meist von Werken der westlichen Kunstmusik (z.B. Kühn, 1987; de la Motte, 1990).

Aber auch Analyseverfahren aus der kognitiven Musikwissenschaft greifen häufig auf das Konzept melodischer Ähnlichkeit zurück. Besonders einflussreich waren für diesen Bereich die Theorien Heinrich Schenkers und die Generative Theory of Tonal Music (Lerdahl & Jackendoff, 1983), die ‚echte‘ melodische Ähnlichkeit und Verwandtschaft als (partielle) Gleichheit auf einem gewissen strukturellen Abstraktionsniveau melodischer Gebilde definierten. Eine ganze Reihe von Theorien und Methoden zur Ähnlichkeitsbestimmung sind mit diesem Ansatz – die strukturelle, tieferliegende Ähnlichkeiten zu ermitteln und im Gegensatz dazu Details der melodischen Oberfläche geringer zu bewerten – seit Anfang der 1980er Jahre entstanden.

Als Beispiel für diesen Ansatz mag die epochengeschichtliche Analyse von melodischen Strukturen mit Hilfe von generativen Grammatiken durch die Gruppe um Mario Baroni (z.B. Baroni & Callegari, 1984; Baroni et al., 1992) gelten. Baroni und Kollegen gehen von einem Kern („kernel“) kurzer melodischer Phrasen aus, der durch die Start- und Endtöne der Phrase gebildet wird. Dieser Kern kann auf verschiedene oder ähnliche Weise durch die Anwendung generativer Einfügingsregeln mit Tönen gefüllt werden. Ähnlichkeit wird im Sinne der Übereinstimmung der Kerntöne, der melodischen Kontur und der Verwendung von Einfügingsregeln definiert. Das Modell von Baroni und Mitarbeitern zeigt eine gewisse Verwandtschaft zu den melodischen Schemata, die Leonard Meyer (z.B. 1973) vorgeschlagen hat und die einen der Pfeiler für Mark Schmucklers Erwartungskonzept bilden (s. Kap. 3.3.3.3). Nicht anhand generativer Regeln, aber mit Hilfe des Vergleiches ‚wichtiger‘ Struktureigenschaften von melodischen Folgen arbeitet das von Eleanor Selfridge-Field entworfene Ähnlichkeitsmaß (Selfridge-Field, 1998, 2002). Ermittelt wird Ähnlichkeit hier mittels einer Bewertungsprozedur, die die Übereinstimmung in verschiedenen Parametern registriert und diese gegeneinander gewichtet verrechnet.

c) Musikwissenschaftliche Beschäftigung mit Urheberrechtsfragen:

Das Konzept der melodischen Ähnlichkeit spielt in der Rechtspraxis im Bereich des musikalischen Urheberrechts eine gewichtige Rolle. Denn den besonderen Melodienschutz und damit die wirtschaftliche Verwertbarkeit sichert das deutsche Urheberrecht (nach §24, Abs. 2) nur den Melodien zu, die neben anderen Voraussetzungen die Forderung erfüllen, nicht aus einem anderen Werk der Musik übernommen zu sein. D.h. ist eine Melodie unzulässigerweise aus einem vorbestehenden Musikwerk übernommen worden, welches sich im Bereich der gesetzlichen Schutzfrist befindet, so handelt es sich bei dem neuen Werk um ein Plagiat, dessen wirtschaftliche Nutzung nicht dem Schöpfer (= Plagiateur) zusteht, sondern – zumindest teilweise – dem ersten Urheber der Melodie.

Sind beide Melodien identisch, so ist allerhöchstens unter der Annahme einer sogenannten zufälligen Doppelschöpfung bestreitbar, dass es sich um eine Melodieübernahme aus dem Originalwerk handelt. Bei bekannten Melodien, die im Kulturraum des Schöpfers der zweiten Melodie weit verbreitet sind oder waren, wird eine zufällige Doppelschöpfung jedoch in der Regel für äußerst unwahrscheinlich gehalten.

Häufiger ergibt sich in der Praxis der Fall, dass beide Melodien zwar nicht identisch sind, beim Hörer jedoch einen gewissen Ähnlichkeitseindruck hinterlassen. Dies lässt i.d.R. die Schöpfer und wirtschaftlich Nutzungsberechtigten der Originalmelodie vermuten, dass bei

der Übernahme der Melodie in ein neues Werk bewusst unwesentliche Teile der Melodie verändert worden sind, um die Übernahme zu kaschieren, die wesentlichen Teile, die den eigenständigen Charakter der Melodie ausmachen und deren wirtschaftlichen Erfolg begründen, aber unverändert übernommen wurden. Nach dieser Argumentation sind sich die Originalmelodie und die Melodie des neuen Werkes hinreichend ähnlich, um von einem Plagiat sprechen zu können. Die Frage der Ähnlichkeit von Melodien ist in diesem Zusammenhang also zentral, und von ihrer Beantwortung hängen nicht selten beträchtliche Summen ab, die aus den Erlösen der mechanischen Vervielfältigung und der öffentlichen Aufführung und Wiedergabe des Werkes stammen.

Die Beurteilung der Ähnlichkeit von Melodien in diesen urheberrechtlichen Streitfragen wird in den meisten Fällen Experten übertragen, in der Regel Musikwissenschaftlern, Musikpraktikern oder Spezialisten aus Verwertungsgesellschaften oder Musikverlagen. Trotz der Häufigkeit, mit der diese urheberrechtlichen Streitfälle auftauchen (schätzungsweise 50 bis 70 Fälle im Jahr in der Bundesrepublik), werden bislang keine objektiven Methoden zur Bestimmung der melodischen Ähnlichkeit bzw. zur Übereinstimmung von wesentlichen melodischen Merkmalen eingesetzt. Das Urteil der zu Rate gezogenen Experten gründet sich dagegen zumeist auf subjektive Einschätzungen, die ggf. mit einzelnen musikanalytischen Details in Teilen gestützt werden. Zu den häufiger angeführten Details gehören die Anzahl der in beiden Melodien übereinstimmenden Tonhöhen oder Intervalle, die Übereinstimmung von Tönen auf den metrischen betonten Zählzeiten, die Übereinstimmung in der Phrasen- und Abschnittsgliederung sowie rhythmische Übereinstimmungen.

Weder ist die Zulässigkeit des Expertenurteils bei der Beantwortung dieser Art von Streitfragen anzuzweifeln – im Gegenteil ist das Urteil eines kompetenten Experten wahrscheinlich das wünschenswerteste Verfahren zur Entscheidungsfindung –, noch die in den meisten Fällen als Belege angeführten analytischen Details und Zusammenhänge. Jedoch verwundert es, dass trotz der großen Anzahl von Fällen in diesem Bereich bislang keine Formalismen zur Bestimmung von Ähnlichkeit vorgeschlagen oder propagiert wurden. Denn zumindest unter dem Aspekt der Nachvollziehbarkeit der Ergebnisse und damit der Objektivität sowie aus dem Gesichtspunkt der Arbeitersparnis erscheint die Verwendung von Formalismen sehr sinnvoll. Anders ist die Praxis beispielsweise bei bautechnischen oder medizinischen Gutachten, die in der Regel eine Expertenmeinung enthalten, die von den Ergebnissen standardisierter und nachvollziehbarer Methoden ausgehen.

Es liegt daher der Verdacht nahe, dass für den Bereich der melodischen Ähnlichkeit die Einführung standardisierter Methoden in gutachterlichen Beurteilungen weder von Seiten der Gutachter noch der Auftraggeber wirklich gewünscht ist. Denn dies könnte einen geringeren Interpretations- und Verhandlungsspielraum sowie höhere Gutachterkosten nach sich ziehen.

Das Fehlen standardisierter Methoden, die gerichtlichen Entscheidungen zu Melodieplagiaten zugrunde liegen könnten, belegt der Überblick, den Charles Cronin über die Behandlung von melodischer Ähnlichkeit im US-amerikanischen Rechtswesen gibt (Cronin, 1998). Er beklagt zudem die zu einseitige Fixierung der US-amerikanischen Gerichte auf die Parameter der Tonhöhe und der Tondauer. Wie Thomas Jörger (1992) plädiert auch er für die Einbeziehung weiterer musikalischer Eigenschaften wie Sound, Tempo oder ‚Style‘, die bei der Wahrnehmung auch von melodischen Gebilden mitbeteiligt seien.

d) Retrieval Systeme für Melodiedatenbanken:

Das Problem melodischer Ähnlichkeit und der Repräsentation von Melodien unter diesem Gesichtspunkt ist wahrscheinlich so alt wie die Projekte zur systematischen Katalogisierung und Archivierung von Musik nach musikalischen Merkmalen. Als Beispiele dienen das *Dictionary of musical themes* von Barlow und Morgenstern (1949), welches Themen aus 10.000 Kompositionen der westlichen Kunstmusik verzeichnet und diese durch einen Themenindex zusammenordnet, welcher nur die Tonhöhen der ersten sechs bis acht Töne enthält. Unterschiede oder Ähnlichkeiten in allen übrigen musikalischen Dimensionen, wie Rhythmus, Klangfarbe oder Harmonik bleiben unberücksichtigt. Analoge Reduktionen nehmen auch andere gedruckte Melodie- und Incipitsammlungen vor, die dadurch implizit die Ähnlichkeit der gesammelten Melodien untereinander bzw. die Ähnlichkeit der vermutlichen Suchanfrage eines Benutzers mit den verzeichneten Melodien auf eine oder wenige Dimensionen beschränken: So sind die Melodien im *National Tune Index* von Keller und Rabson (1980) auf die Abfolge der Tonleiterstufen und die Intervallfolge beschränkt und das *Directory of tunes and musical themes* von Parsons (1975) sogar nur auf die melodische Kontur.

Mit dem Einsatz des Computers zur Archivierung, Verwaltung und Suche von Melodien ist das Problem der Repräsentation von musikalisch-melodisch Gebilden und der Definition von Ähnlichkeitsfunktionen für die Suche und die Gruppierung in solchen Melodiedatenbanken gewachsen. Zum einen kann durch den praktisch fast unbeschränkten Speicherplatz und die teilweise Automatisierung bei der Erfassung von Melodien die Zahl der

verzeichneten Melodieeinträge leicht auf ein vielfaches der in gedruckten Indizes archivierte Melodien steigen. So sind bei der Verwendung einer wenig differenzierenden Ähnlichkeitsfunktion in vielen Fällen zu große Treffermengen zu erwarten, die die Verwendbarkeit einer solchen Datenbank einschränken. Zum anderen ermöglicht der Einsatz des Computers, Melodiendatenbanken einem viel breiteren Benutzerkreis zur Verfügung zu stellen. In den sogenannten ‚neuen‘ Medien (vgl. Hemker & Müllensiefen, 1996), wie Internet, CD-ROMs oder mobilen Datennetzen ist die Integration von Melodiedatenbanken schon realisiert (s.u.) oder in Planung. Dieser neue Nutzerkreis verlangt nicht nur nach neuen Schnittstellen für Suchanfragen, die einfach und ohne (musikalisches) Vorwissen zu bedienen sind, sondern erwartet auch vom Design dieser Schnittstellen eine höhere Fehlertoleranz bzw. eine Anpassung an die Eigenheiten von ‚laienhaften‘ Melodiesuchen. Zu diesen neuen Schnittstellen zählen das Singen oder Summen von Melodien, die Eingabe von Audiodatenformaten wie *.wav* oder *.mp3* oder die Verwendung einer virtuellen Klaviatur. Als Beispiele für Melodiedatenbanken mit Internetschnittstellen können die Meldex Datenbank der Universität von Waikato (Neu Seeland; URL: <http://www.nzdl.org/musiclib/>), die Themefinder-Datenbank des Center of Computer Assisted Research an der Universität Stanford (URL: <http://www.themefinder.org>) und das *Query-by-humming*-Interface auf dem deutschen Musikportal musicline.de (<http://www.musicline.de/de/melodiesuche>) gelten.

Die Verwendung einer Ähnlichkeitsfunktion, die zu ‚laienhaften‘ Suchanfragen verwertbare Ergebnisse für den Benutzer liefert, ist in allen diesen Schnittstellen von zentraler Bedeutung. In den bereits realisierten Systemen dieses Bereiches werden eine Reihe von verschiedenen Ähnlichkeitsmodellen z.T. alleine, z.T. in Kombination miteinander verwendet. Zu den häufigsten Ähnlichkeitsmaßen zählen die sogenannte Levenstein-Distance (auch Edit Distance) für Tonhöhenwertereihen, der Vergleich von Tonhöhenkonturen und der Vergleich von Wahrscheinlichkeiten über sogenannte *Hidden Markov Models* für Tonhöhen- und Tondauernreihen.

e) Musikpsychologie:

Als Disziplin der musikalischen Grundlagenforschung hat die Musikpsychologie ein quasi ‚natürliches‘ Interesse daran, die Frage zu beantworten, wie die Wahrnehmung von melodischer Ähnlichkeit funktioniert und von welchen musikstrukturellen und subjektabhängigen Eigenschaften sie bestimmt ist.

Ein starkes Interesse leitet sich beispielsweise aus psychologischen Erklärungsversuchen für die melodische Variantenbildung in der Volksliedforschung ab. So bilanziert z.B. schon Walter Wiora in seinem Aufsatz zur Systematik des Umsingens: „Das wirkliche Umsingen geht aber im allgemeinen nur naheliegende Wege“ (Wiora, 1941, S. 131). Obwohl Wiora die Beschreibung dieser ‚naheliegenden Wege‘, d.h. der psychologischen Gesetze für melodische Variantenbildung für möglich hält, will (oder kann) er sie jedoch (zu jenem Zeitpunkt) nicht vornehmen.

Auch für die psychologische Validierung von musikanalytischen Verfahren z.B. zur Analyse nicht-tonaler Musik kann die Klärung der Frage, was Hörer als ähnlich beurteilen, von großer Wichtigkeit sein. Mark Schmuckler vergleicht in diesem Zusammenhang die Ergebnisse zweier musikanalytischer Prozeduren für Werke Neuer Musik mit den Wahrnehmungen von Versuchspersonen und einer weiteren Methode zur Messung von Konturähnlichkeiten, die auf der Fouriertransformation beruht (Schmuckler, 1999).

Der Anspruch des im Prinzip musikpsychologischen Ansatzes von Hofmann-Engl (2001; 2002) ist es, Ähnlichkeit von melodischen Gebilden grundsätzlich zu definieren, sodass sie sowohl für Retrieval-Anwendungen wie auch musikanalytisch Gültigkeit beanspruchen kann.

Einen kursorischen Überblick über die neuere musikpsychologische Forschung zur musikalischen bzw. melodischen Ähnlichkeit geben McAdams und Matzkin (2001). Sie fassen hier auch zwei eigene Experimente zusammen, auf deren Grundlage sie ein einfaches additives Modell entwickeln, das die wahrgenommene Ähnlichkeit durch die Anzahl der Abweichungen zwischen zwei Melodien in den einzelnen Parametern zu erklären versucht.

Wie sehr ein Modell melodischer Ähnlichkeit für den Bereich der musikalischen Gedächtnisforschung, der selbstverständlich auch zur Musikpsychologie zu zählen ist, von Nöten ist, soll die folgende etwas ausführliche Darlegung deutlich machen.

4.1.2 Definition melodischer Ähnlichkeit als Grundlage für die Modelle des Melodiegedächtnisses

Auch für den in dieser Arbeit interessierenden Aspekt des Melodiegedächtnisses ist das Konzept der melodischen Ähnlichkeit zentral. Im Verlauf dieser Arbeit wird ein Konzept von melodischer Ähnlichkeit sowohl für die Analyse von Gedächtnisrepräsentationen von Melodien im Experiment wie auch für die Formulierung allgemeiner Gedächtnisregeln

eingesetzt werden. Schließlich enthält auch der praktische Anwendungsbezug dieser Arbeit – das *Query-by-humming*-System – in seinem Kern und ganz unabhängig von Gedächtniseffekten einen Zuordnungsmechanismus, der über eine melodische Ähnlichkeitsfunktion definiert ist.

Die Analyse von melodischen Gedächtnisrepräsentationen ist in der Regel notwendiger Bestandteil der Planung und Auswertung von Gedächtnisexperimenten mit musikalischen Materialien, und zwar sowohl im Wiedergabe- (*recall*) wie im Wiedererkennensparadigma (*recognition*).

Beispielsweise kamen Sloboda & Parker (1985) bei der Bewertung ihres Wiedergabeexperiment zu dem Schluß, dass die Wiedergaben ihrer Versuchspersonen zwar fehlerhaft im Vergleich mit der Originalmelodie sind, aber dennoch „highly related to the original in many respects“ (S. 159). Gerade die Struktur der Testmelodien werde gut behalten, Oberflächendetails würden dagegen eher abweichen. Zu diesen Schlüssen gelangen sie hauptsächlich aufgrund von Fehlerauszählungen bei den wiedergegebenen Melodievarianten und der Bewertung durch ihren Sachverstand. Die beiden Forscher bemängelten in dem selben Artikel von 1985 jedoch, dass keine psychologische Theorie der melodischen Identität existiere, die diese strukturellen Übereinstimmungen objektivieren könne. Zur Frage, ob einzelne Abweichungen Oberflächendetails oder strukturtragende Elemente betreffen, konnten Sloboda und Parker also nur begründete Mutmaßungen anstellen, sie aber nicht mit Hilfe eines eindeutigen Formalismus beantworten. Da sich ein zukünftiges Gedächtnisexperiment im Wiedergabeparadigma auf die Analyse von menschlich generierten Varianten zu vorgegebenen Melodien stützen soll, ist auch hierfür die Frage nach der Ähnlichkeit von Original und Variante zuerst zu objektivieren. Die Ähnlichkeit einer wiedergegebenen Variante dient dabei als abhängige Variable. Nur wenn diese Variable objektiv mess- und skalierbar ist, lassen sich in einem Wiedergabeexperiment die Einflüsse von verschiedenen unabhängigen subjekt- (z.B. musikalische Expertise) und designbezogenen (z.B. Anzahl der Wiederholungen) sowie melodisch-strukturellen (z.B. Anzahl Motivwiederholungen) Variablen auf das melodische Gedächtnis präzise feststellen.

Auch in Experimenten, die das Wiedererkennensparadigma für das Wiedererkennen echter melodischer Passagen verwenden (z.B. Kauffman & Carlsen, 1989; Dowling et al., 2002), kann die melodische Ähnlichkeit von Original und Variante eine entscheidende Rolle spielen. Hier tritt sie als unabhängige Variable auf, über die die Ähnlichkeit der zu vergleichenden Versuchsstimuli beschrieben und kontrolliert wird. Sollen beispielsweise die Ef-

fekte bestimmter melodisch-struktureller Eigenschaften auf die Wiedererkennensleistung untersucht werden, so ist es geboten, die Stimuli nur entlang dieser Variable zu variieren, die globale Ähnlichkeit zwischen Original und Variante aber konstant zu halten. Auch hierfür ist eine genaue Bestimmung melodischer Ähnlichkeiten unumgänglich.

Ziel dieser Arbeit ist die Formulierung von Regeln, die Fehler bzw. Konstanzen vorhersagen, durch die sich Transformationen von Melodien im Gedächtnis beschreiben lassen, kurz die Modellierung der Mechanismen des Melodiegedächtnisses. Als Basis für diese Modellierung dient eine Ähnlichkeitsfunktion, die auf Grund determinierender Größen wie z.B. struktureller Bau der Originalmelodie, musikalische Expertise der Versuchspersonen, Höraufmerksamkeit oder Anzahl der Wiederholungen einer Melodie mit einer gewissen Irrtumswahrscheinlichkeit den Grad der melodischen Ähnlichkeit angibt, die die resultierende Gedächtnisrepräsentation einer Melodie bei einer Vpn. aufweist. Zusätzlich zu diesem Ähnlichkeitsgrad sind die Typen von Abweichungen oder Fehler der Gedächtnisrepräsentation vom Original zu spezifizieren und zu quantifizieren. Diese Art von Vorhersagen gehen zwar über die reine Vorhersage eines Ähnlichkeitsgrades hinaus, lassen sich aber anhand derselben Parameter beschreiben (z.B. Intervallfolge, Kontur, Rhythmus, Tonalität), die auch für die Konstruktion von Ähnlichkeitsmaßen maßgeblich sind. Die Erkenntnisse bezüglich Konzeptionierung, Skalierung und Messung dieser Parameter, welche bei der Erstellung des Ähnlichkeitsmaßes geleistet werden müssen, lassen sich später ggf. auf die Beschreibung von Gedächtnis(fehler)mechanismen übertragen.

Bei der Suche nach geeigneten Ähnlichkeitsmodellen für Melodien zeigt sich schnell, dass das Problem nicht im Mangel an existierenden Ähnlichkeitsbeschreibungen besteht, sondern im Gegenteil in deren Überfülle. In jedem der fünf oben aufgezählten Bereiche sind bereits mehrere Methoden zur Bestimmung melodischer Ähnlichkeit vorgeschlagen worden.

Es drängt sich deshalb sofort die Frage auf, welches Ähnlichkeitsmodell für den hier interessierenden Gedächtnisbereich am gewinnbringendsten eingesetzt werden kann. Hier kommen prinzipiell viele der oben aufgezählten Ähnlichkeitsmodelle in Frage, die in der Regel den Anspruch erheben, Ähnlichkeit zwischen Melodien nicht nur in einem bestimmten Bereich, sondern ganz allgemein bestimmen zu können. Somit ist ein Kriterium von Nöten, nach dem die Angemessenheit der verschiedenen Modelle beurteilt werden kann.

Am naheliegendsten ist die Idee, das Modell auszuwählen, welches die ‚besten‘ Ergebnisse bei der Analyse der nachfolgenden Gedächtnistests liefert. ‚Gute‘ bzw. ‚bessere‘ Ergebnisse könnten im Sinne von ‚differenziert‘, ‚robust‘, ‚trennscharf‘ oder ‚effektiv‘ operationalisiert werden. Da jedoch im Vorfeld der Gedächtnisexperimente nichts über die Ausprägung der zu erwartenden Ergebnisse gesagt werden kann, bzw. das Auffinden dieser Ergebnisse ja gerade das Ziel der Experimente ist, verbietet die wissenschaftliche Logik, die Ergebnisse der Experimente selbst als Gütekriterium eines Analyseinstruments derselben Ergebnisse zu verwenden.

Da es also schwierig werden dürfte, ein objektives Kriterium für die Angemessenheit eines Ähnlichkeitsmodells für den vorliegenden analytischen Zweck zu finden, ohne in einen definitorischen Zirkel zu geraten, bietet sich ein Vorgehen an, welches neben einer zirkellosen Argumentation auch noch den Vorteil besitzt, Erkenntnisse über das menschliche Urteilsverhalten hinsichtlich melodischer Ähnlichkeiten als ‚Nebenprodukt‘ zu liefern: Der Vergleich der verschiedenen Ähnlichkeitsmodelle mit den Bewertungen kompetenter menschlicher Beurteiler. Dieser Ansatz ‚objektive‘, messbare melodische Ähnlichkeit mit ‚subjektiver‘, von Menschen wahrgenommener Ähnlichkeit zu vergleichen, wurde zwar von einigen Autoren von Zeit zu Zeit eingefordert (vgl. z.B. Jesser, 1990, S. 20), doch bislang wurden selten konkrete Versuche in dieser Richtung unternommen (Ausnahmen sind z.B.: Egmond et al., 1996; Hofmann-Engl, 2002). Die Vorteile, die Übereinstimmung mit subjektiven Ähnlichkeitsurteilen als Kriterium für die Angemessenheit von Ähnlichkeitsmodellen zu wählen, liegen auf der Hand:

- Die Ergebnisse einer Ähnlichkeitsbestimmung, die der menschlichen Wahrnehmung von Ähnlichkeit nahe ist, sind – auch für musikalische Laien – gut nachvollziehbar.
- Kompetente Beurteiler dienen als Referenzpunkt und verleihen dem Verfahren so Autorität. Das Einholen von Expertenmeinungen ist in vielen Streitfragen und komplexen Sachgebieten eine gängige Methode.
- Da die Ähnlichkeitsbewertungen der menschlichen Urteiler empirisch gewonnen werden, ist eine Wechselwirkung mit den zum Vergleich heranzuziehenden Modellen ausgeschlossen. D.h. alle Modelle müssen sich auf gleiche Weise mit den Bewertungen der Versuchspersonen messen lassen.
- Ein argumentativer Zirkelschluß wird durch den Bezug auf eine äußere Referenz vermieden.

- Ist die Übereinstimmung zwischen einem Ähnlichkeitsmodell und den menschlichen Urteilen hinreichend gut, so lässt sich dieses Modell auch als Beschreibung des menschlichen Urteilsprozesses interpretieren. Die analytische Betrachtung dieses Ähnlichkeitsmodells lässt dann Rückschlüsse auf die Komponenten des menschlichen Ähnlichkeitsurteils zu.
- Liegen empirische Daten zur Wahrnehmung einmal vor, so können bestehende Modelle in Richtung dieser Daten optimiert und erweitert werden; d.h. es ist möglich, ein aussichtsreiches Modell im Nachhinein an diese Daten ‚anzufitten‘.

Es soll also dasjenige Ähnlichkeitsmodell zur Analyse von Gedächtnisfehlern und zur Konstruktion von Regeln für die Beschreibung von Gedächtnisfehlern verwendet werden, welches die größte Übereinstimmung mit den empirisch gewonnenen Urteilen von kompetenten Versuchspersonen zeigt.

Im folgenden sollen zunächst einige der erfolgversprechendsten Ähnlichkeitsmodelle kurz beschrieben und für eine Verwendung im nachfolgenden Experiment operationalisiert werden. Dabei ist jedes Modell als Stellvertreter für eine Klasse von ähnlichen Modellen zu verstehen. Das Spektrum der ‚auf dem Markt‘ befindlichen Modelle soll so zumindest teilweise ausgeleuchtet und auf seine Übereinstimmung mit dem menschlichen Urteil geprüft werden.

Danach wird das Vorgehen zur Konstruktion des Versuchsdesigns und der Versuchsstimuli sowie die rechnerische Behandlung der empirischen und der gemessenen ‚objektiven‘ Ähnlichkeitsdaten geschildert. Der Abschnitt *Ergebnisse* zeigt die Beziehungen zwischen den menschlichen Ähnlichkeitsbeurteilungen und den verschiedenen objektiven Verfahren der Ähnlichkeitsbestimmung auf und betrachtet sie im Kontext der an der Melodiekonstruktion beteiligten Variablen.

4.2 Die zu testenden Ähnlichkeitsmodelle

4.2.1 Vorstellung der Modelle

Bei dieser Vorstellung stehen weniger die genaue Herleitung und die erkenntnistheoretische Einbettung der verschiedenen Ähnlichkeitsmodelle in die verschiedenen Forschungsansätze oder Wahrnehmungstheorien im Zentrum des Interesses als die Beschreibung der

Operationalisierungen, durch die die Modelle im Kontext dieser Arbeit verwendbar werden. Es handelt sich bei allen Modellen mit Ausnahme des Modells von Selfridge-Field um (z.T. einfache) mathematische Modelle, die in der Regel nur einen musikalischen Parameter von Melodien numerisch erfassen und vergleichen können.

Die Konstruktion der Modelle selbst auf Grundlage von empirischen Erkenntnissen, musikalischer Sachkenntnis und logischen Schlussfolgerungen kann in diesem Rahmen nur am Rande diskutiert werden. Für eine tiefergehende Diskussion wird auf die Originalliteratur verwiesen. Wichtigstes Kriterium für die Angemessenheit der Modelle soll in dieser Arbeit ihre Vorhersagequalität in Hinsicht auf die empirisch gewonnenen menschlichen Beurteilungen sein.

Zur Illustration der verschiedenen Modelle wird im folgenden immer dasselbe Beispiel zweier ähnlicher Melodievarianten verwendet. Das Beispiel entstammt der Studie von Sloboda und Parker (1985) und ist dort als Transkription im Notenbild dokumentiert (Sloboda & Parker, 1985, S. 165; s. Abb. 4-1).



Abb. 4-1: Testmelodie aus Sloboda & Parker (1985)

Es ist die vierte gesungene Wiedergabe der siebten Versuchsperson. Als Originalmelodie diente ein russisches Volkslied, welches den Versuchspersonen unbekannt war. Diese Wiedergabe 7.4 eignet sich als Beispiel besonders gut, da sie in etwa die gleiche Länge wie das Original aufweist und auch darüber hinaus auf den ersten Blick viele musikalische Charakteristika mit ihrem Original gemein hat, sodass eine Ähnlichkeitsbestimmung sinnvoll durchführbar erscheint. Außerdem stammt die Beispielvariante von einer der Versuchspersonen von Sloboda und Parker. Die Abweichungen vom Original in den verschiedenen musikalischen Parametern sind also quasi natürlich und dürften durch Intonationsfehler und in stärkerem Maße durch eine fehlerhafte Gedächtnisrepräsentation zustande gekommen sein. Gerade in Hinblick auf die in dieser Arbeit interessierende gedächtnispsychologische Fragestellung ist also der Vergleich der von Sloboda und Parker dokumentierten Gesangswiedergabe mit ihrem Original interessant.

Aus Gründen der Vergleichbarkeit untereinander und der universellen Verwendbarkeit sollen alle der für die eigene Arbeit operationalisierten Ähnlichkeitsmaße folgende Eigenschaften besitzen:

Die Ähnlichkeitsmaße sollen die Ähnlichkeit durch einen numerischen Wert aus dem Bereich von 0 bis 1 ausdrücken, wobei 0 für die größtmögliche Unähnlichkeit und 1 für die größte Ähnlichkeit bzw. Identität der beiden zu vergleichenden Melodien steht.

Die Ähnlichkeit wird als symmetrische Relation angenommen, d.h. der Ähnlichkeitswert für den Vergleich von Melodie A mit B wie von Melodie B mit A ist identisch.

Die melodische Ähnlichkeit berücksichtigt nicht die absolute Tonhöhe und das absolute Tempo der Melodien, d.h. sie ist transpositions- und dilatationsinvariant in den Dimensionen Tonhöhe bzw. Tempo.

Die in der Literatur beschriebenen Originalversionen der meisten Maße besitzen diese Eigenschaften nicht. Die notwendigen mathematischen Transformationen wurden zusammen mit Klaus Frieler im Rahmen eines informellen, aber sehr ertragreichen Forschungsprojektes zur Ähnlichkeitsanalyse von Melodien am Musikwissenschaftlichen Institut der Universität Hamburg gemeinsam entworfen.

A) Melodien als zeitabhängige Tonhöhenverläufe:

Neben den statistischen Merkmalen, wie den Häufigkeitsmerkmalen von Tonhöhen und Tondauern, die Wolfram Steinbeck (1982) für die clusteranalytische Einteilung seines Melodiebestandes benötigt, schlägt er auch Ähnlichkeitsbeziehungen vor, die melodische Eigenschaften in den Dimensionen Zeit und Reihenfolge verwenden (S. 102-131).

Das einfachste dieser Modelle nennt Steinbeck „Treppenkurve“. Die Treppenkurve stellt eine Übertragung der Dimensionen Tonhöhe und Dauer in das kartesische Koordinatensystem dar.

Die Tonhöhenstufen werden dabei numerisch kodiert und die Dauern als Vielfache einer kleinsten Dauerneinheit dargestellt. Die folgende Abbildung macht dieses Prinzip deutlich:

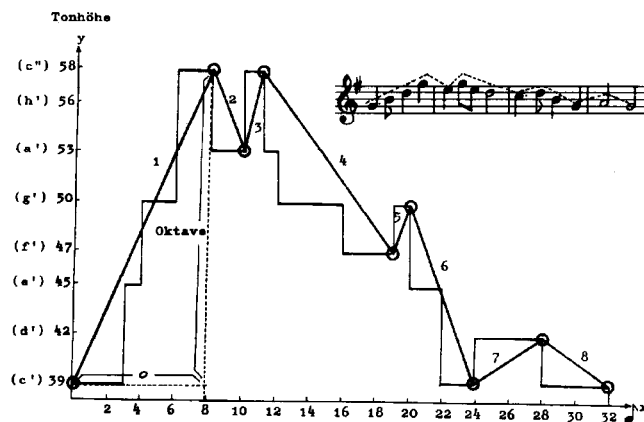


Abb. 4-2: Abbildung 44 aus Steinbeck (1982, S. 111).

Für den Anfang der von Sloboda und Parker verwendeten Originalmelodie lässt sich das Notenbild von Abb. 4-1 in die numerische Kodierung der Tonhöhen im MIDI-System folgendermaßen übertragen:

Note	a	g	f	g	f	e	e	a	a	f	d	d	a	g	f	g	f	e	a	d	c	c ²	c ²	a	a	b	b	a	g	a
MIDI-Nr.	69	67	65	67	65	64	64	69	69	65	62	62	69	67	65	67	65	64	69	62	60	72	72	69	69	70	70	69	67	69

Die Dauern der einzelnen Noten sind in dieser Darstellung nicht mitkodiert. Steinbeck verwendet für die numerische Kodierung der Tonhöhen das Melodiencodesystem MECOS, das auch enharmonische Unterschiede darstellen kann. Die numerische Darstellung von Tonhöhen wird in dieser Arbeit ausschließlich mit Hilfe des weit verbreiteten MIDI-Standards geschehen, der eine große Anzahl von Vorteilen für die praktische Arbeit aufweist.

Zur Berücksichtigung der Dauern vervielfacht Steinbeck jede Tonhöhe um das Vielfache des Dauernwertes im Vergleich mit dem kürzesten in der Melodie vorkommenden Dauernwert. Dies ist in der Beispielmelodie die 8tel Note. So ergibt sich die folgende um die Dauern gewichtete Tonhöhenreihe:

Note	a	g	f	g	f	e	e	a	a	f	d	d	a	g	f	g	f
MIDI-Nr. Gew.	69 69 69	67	65 65	67	65	64	64	69	69	65	62 62	62	69 69 69	67	65 65	67	65

Note	e	a	d	c	c ²	c ²	a	a	b	b	a	g	a
MIDI-Nr. Gew.	64 64	69 69	62 62 62 62	60 60	72	72	69 69 69	69	70	70	69	67	69 69

In dieser Tonhöhenreihe wird nun zwar nicht mehr zwischen Toneinsätzen bzw. repetierten Tönen und ausklingenden Tönen sowie Pausen unterschieden, dafür hat diese Darstellung

jedoch den Vorteil, dass Töne, die länger erklingen, stärker gewichtet werden, was wahrnehmungspsychologisch beispielsweise durch die Untersuchungen von Povel und Essens gestützt wird (Povel & Essens, 1985).

Aufgrund von technischen Schwierigkeiten bei der Extraktion der Tondauern aus MIDI-files, wird in der Implementation aller Ähnlichkeitsmodelle, die auf dieser numerischen Darstellung beruhen, die Dauer des letzten Tones immer bis zum nächsten Taktviertel ergänzt. Endet der letzte Ton auf einem Taktviertel, so bekommt er die Länge von einem Viertel, in diesem Beispiel also zwei Achtel.

Es ist nach diesem Kodierungsverfahren also möglich, die Tonhöhen- und Dauernwerte einer Melodie auf plausible Weise in eine Reihe numerischer Werte zu transformieren.

Für zwei Melodien, die miteinander verglichen werden sollen, lassen sich analog zwei Wertereihen bilden, wie anhand der folgenden Variante 7.4 der Originalmelodie nachzuvollziehen ist:



Abb. 4-3: Vierte Wiedergabe der Testmelodie durch Vpn. 7 im Versuch von Sloboda & Parker (1985)

Numerische Darstellung (ungewichtet):

Note	a	g	f	e	f	f	a	f	d	a	g	f	d	d	e	a	f	d	d	d2	c2	c2	b	b	a	a	f	g
MIDI-Nr.	69	67	65	64	65	65	69	65	62	69	67	65	62	62	64	69	65	62	62	74	72	72	70	70	69	69	65	67

Numerische Darstellung (gewichtet):

Note	a	g	f	e	f	f	a	f	d	a	g	f	d	d	e	a
MIDI-Nr. gew	69 69 69	67	65 65	64	65	65 65 65 65	69 69	67 67	62 62 62 62	69 69 69	67	65 65	62	62	64 64	69 69

Note	f	d	d	d2	c2	c2	b	b	a	a	f	g
MIDI-Nr. Gew.	65 65 65 65 65 65 65 65	62	62 62	74	72	72	70	70	69 69	69 69	65	67

Als Ähnlichkeitsmaße für den Vergleich zweier solcher Wertereihen schlägt Steinbeck die durchschnittliche Differenz zwischen den beiden Wertereihen und den Produkt-Moment-Korrelationskoeffizienten nach Pearson-Bravais vor.

Die durchschnittliche Differenz ist die Summe der Differenzbeträge zwischen allen Tonhöhenwerten, und sie ist bei Steinbeck folgendermaßen definiert:

$$d = \frac{\sum |y_i - x_i|}{n}$$

Dabei ist y_i ein Ton der ersten Melodie an Position i und x_i ein Ton der zweiten Melodie an Position i . Die durchschnittliche Differenz wird im Nenner an der Anzahl der verglichenen Tonpaare normiert. Das Maß d hat jedoch zwei Nachteile: Erstens fehlt eine Normierung, durch die es auf den Wertebereich zwischen 0 und 1 skaliert werden kann. Zweitens ist die durchschnittliche Differenz gegenüber der Transposition einer Melodie in eine andere Tonart nicht invariant. Da die Transpositionsinvarianz jedoch ein unerlässliches Kriterium für die hier verwendeten Ähnlichkeitsmaße ist, werden zwei Verfahren genutzt, die diese Transpositionsinvarianz garantieren.

Das erste Verfahren schaltet der Berechnung von d die Transposition einer Melodie auf die Tonhöhe der anderen Melodie vor. D.h. es werden beide Melodien auf die gleiche Tonhöhe gebracht. Dies geschieht durch folgenden, in Zusammenarbeit mit Klaus Frieler (vgl. Müllensiefen & Frieler, 2004) entwickelten Algorithmus:

- Es werden die Differenzen zwischen den jeweils tiefsten und höchsten Tönen, den häufigsten Tönen und den mittleren Tonhöhen der beiden zu vergleichenden Melodien gebildet.
- Die größte dieser Differenzen wird als Transpositionsdivergenz definiert.
- Die zweite der beiden Vergleichsmelodien wird entlang der Transpositionsdivergenz halbtöne in Richtung der ersten Melodie transponiert.
- Für jeden Transpositionsschritt wird die Ähnlichkeit der beiden Melodien nach der Formel $\frac{rawed + rawedw}{2}$ berechnet. Dabei stehen $rawed$ und $rawedw$ für das ungewichtete bzw. das gewichtete Ähnlichkeitsmaß der Edit Distance, welches weiter unten erläutert und definiert wird.
- Als richtiger Transpositionsschritt wird derjenige definiert, der die größte Ähnlichkeit zwischen den beiden Melodien ergibt (den größten Wert für $\frac{rawed + rawedw}{2}$ liefert)

und bei dem die beiden mittleren Tonhöhen nicht mehr als 2,5 Halbtöne auseinanderliegen.

Dieser Algorithmus geht also davon aus, dass die Edit Distance in ihrer gewichteten und ungewichteten Form bereits eine gute Näherung der ‚wahren‘ Ähnlichkeit liefert, die als Heuristik für die Angemessenheit einer Transposition auf einer gewissen Stufe dienen kann.

Das zweite Verfahren verwendet statt der absoluten Tonhöhen die Abstände (Intervalle) zwischen beiden Melodien als Eingangsdaten, die miteinander verglichen werden. Die durchschnittliche Abweichung wird also als Summe aller Abweichungen jedes Intervalls zwischen zwei Tönen aus beiden Melodien von seinem jeweiligen Vorgängerintervall definiert:

$$diff = 1 - \frac{\sum ((y_{i+1} - x_{i+1}) - (y_i - x_i))}{n \cdot \max(|y_{i+1} - y_i| + |x_{i+1} - x_i|)}$$

Mit y_i als Ton der ersten Melodie an Position i und x_i als Ton der zweiten Melodie an Position i . Die Summe aller Abweichungen wird im Nenner an der Anzahl der verglichenen Intervalle n und der maximalen Summe der Abstände zwischen den aufeinanderfolgenden Tönen der beiden Melodien normiert. Durch diese Normierung wird die Größe der Sprünge berücksichtigt, die in beiden Melodien vorkommen. Melodien, die viele große Intervalle beinhalten und dadurch eine große absolute durchschnittliche Differenz (Zähler des Bruches) aufweisen, sollen einen größeren Ähnlichkeitswert bekommen als Melodien, die nur kleine Sprünge aufweisen, aber dieselbe absolute durchschnittliche Differenz (Zähler) haben. $diff$ kann sich zwischen 0 und 1 bewegen.

Bei der Verwendung der verschiedenen Maße der durchschnittlichen Differenz sowie bei dem nachfolgend erklärten Korrelationskoeffizienten ist für den Vergleich von Werten, die die Verläufe von echten Melodien wiedergeben, an einer Stelle Vorsicht geboten:

Da die durchschnittliche Differenz und der Korrelationskoeffizient immer nur korrespondierende Wertepaare miteinander vergleichen, kann ein zusätzlicher Auftakt bei nur einer Melodie eine ‚Verschiebung‘ einer Wertereihe und so einen deutlich geringeren Ähnlichkeitswert zur Folge haben. Das optionale Vorkommen von Auftakten ist aber bei Melodievarianten z.B. in großen Volksliedsammlungen ein sehr häufiges Phänomen. In der menschlichen Wahrnehmung spielen Auftakte jedoch in der Regel eine geringe Rolle. Es

ist deshalb sinnvoll, wie es beispielsweise Leppig vorschlägt (Leppig, 1987a), die zu vergleichenden Melodien um ein bis mehrere Dauerseinheiten gegeneinander zu verschieben und jeweils den Korrelationskoeffizienten zu berechnen. Der größte dieser so gewonnenen Koeffizienten gilt dann als der ‚wahre‘ Ähnlichkeitswert. In der Implementierung für diese Arbeit wird die kürzere Melodie entlang der längeren verschoben, und zwar von acht Einheiten der kleinsten Dauer links des Beginns bis acht Einheiten oder höchstens 10% der Länge der kürzeren rechts vom Beginn der längeren Melodie. Die zu vergleichenden Melodien werden auf die selbe Anzahl an Dauerseinheiten begrenzt, d.h. die eventuell überzähligen Dauerseinheiten am Ende der längeren Melodien werden vor Berechnung des Korrelationskoeffizienten ‚abgeschnitten‘. Da die Verschiebung nur innerhalb der definierten Grenzen geschieht, ist sichergestellt, dass die durchschnittliche Differenz und der Korrelationskoeffizient als Ähnlichkeitswert immer noch die Ähnlichkeit beider Melodien als Ganzes wiedergibt und nicht beispielsweise die Übereinstimmung zweier Töne am Ende der einen und am Anfang der anderen als hohe Globalähnlichkeit messen kann.

Der Produkt-Moment-Korrelationskoeffizient r nach Pearson-Bravais gibt im Gegensatz zur durchschnittlichen Differenz die Gleichförmigkeit der relativen Bewegung der beiden Melodienverläufe an, wie Steinbeck erklärt (S. 109). Der Korrelationskoeffizient hat also gegenüber der durchschnittlichen Differenz den Vorteil, dass er gegenüber musikalischen Transpositionen an sich schon invariant ist. Er kann Werte zwischen -1 (maximale Gegenläufigkeit) und 1 (maximale Ähnlichkeit) annehmen. $r = 0$ wäre als maximale Unähnlichkeit zu deuten. Bekanntermaßen errechnet sich der Korrelationskoeffizient wie folgt:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Die Gegenläufigkeit (Wertebereich -1 bis 0) kann im Rahmen dieser Arbeit allerdings kein plausibles Kriterium für die Ähnlichkeit von Melodien und Varianten sein, denn es gibt keine Indizien dafür, dass melodische Gedächtnisrepräsentationen als Spiegelung an der horizontalen Achse (=Umkehrung) der Originalmelodien organisiert wären. So wurde in der vorliegenden Untersuchung neben dem von -1 bis 1 skalierten Korrelationskoeffizienten eine ‚geschrumpfte‘ Variante betrachtet, bei der alle negativen Korrelationswerte $= 0$ gesetzt wurden. Diese Variante wird im weiteren Verlauf als r_{st} (‚st‘ steht für ‚stretched‘, obwohl es sich mathematisch gesehen nicht um eine Dehnung des Wertebereichs handelt,

sondern um eine Verkürzung) oder *rawpcst* bezeichnet („raw“ steht für die „rohen“ Tonhöhendaten, die als Eingangsdaten verwendet werden; „pc“ steht für die Pearson-Bravais-Korrelation und „st“ steht wiederum für „stretched“).

Für das obige Beispiel der Testmelodie und ihrer im vierten Versuch von Vpn. 7 gesungenen Variante ergeben sich somit folgende Ähnlichkeitswerte:

Ähnlichkeitswerte ungewichtet:

Produkt-Moment-Korrelation: $r = \text{rawpcst} = 0,667$

(Dies ist der maximale Wert aller nach den Verschiebungen berechneten r ; der maximale Wert kam bei einer Verschiebung der Variante 7.4 um zwei Töne nach rechts zustande;).

Da die durchschnittliche Differenz in der ungewichteten Form wenig sinnvoll erscheint und auch Steinbeck der gewichteten Variante den Vorzug gibt, ist auf eine Implementierung der ungewichteten Variante verzichtet worden.

Ähnlichkeitswerte mit gewichteten Tondauern:

Durchschnittliche Differenz: $\text{diff} = 0,352$

Produkt-Moment-Korrelation: $r = \text{rawpcstw} = 0,47$

Die Ähnlichkeitswerte an sich haben jedoch keine absolute Bedeutung. Wie Steinbeck nachdrücklich unterstreicht, gewinnen die Ähnlichkeitswerte erst dann an Anschaulichkeit und Bedeutung, wenn man sie mit den Ähnlichkeitswerten anderer Melodiepaare vergleicht (Steinbeck, 1982, S. 108-109). Ähnlichkeit ist also immer relativ im Sinne „ähnlicher als“ oder „ähnlich im Vergleich mit“ zu verstehen. Ähnlichkeit ist, wie es Kluge in seinem 2. Relativitätssatz der Ähnlichkeit formuliert, „immer vom Objekthorizont der Betrachter abhängig“ (Kluge, 1996, S. 94). Als Vergleichsobjekte kommen z.B. weitere Varianten desselben Liedes in Frage oder die Melodien einer Melodiedatenbank. Erst Aussagen, wie „die Variante 7.4 ist der originalen Testmelodie von Sloboda und Parker ähnlicher als 98,5% der Melodien einer großen Volkslieddatenbank“ oder „von den sechs Wiedergaben der Versuchsperson 7 ist Variante 7.4 dem Original am ähnlichsten“, geben der Verwendung von numerischen Ähnlichkeitswerten Sinn.

Als Maße für den Ähnlichkeitsvergleich sollen die folgenden Verfahren in dieser Arbeit verwendet werden:

<i>diff</i>	durchschnittliche Differenz der Intervallabstände zwischen beiden Melodien
<i>rawpc</i>	Korrelationskoeffizient nach Pearson-Bravais der Tonhöhenwerte
<i>rawpcw</i>	gewichteter Korrelationskoeffizient nach Pearson-Bravais der Tonhöhenwerte
<i>rawpcst</i>	auf den Wertebereich 0-1 transformiertes r durch Setzen der negativen Werte gleich 0
<i>rawpcwst</i>	auf den Wertebereich 0-1 transformiertes r durch Setzen der negativen Werte gleich 0, zusätzlich gewichtet

B) Melodien als Verläufe der Tonhöhenkontur:

Bereits im Kapitel 3.2 *Gedächtnis für kurze Tonfolgen* wurden zahlreiche empirische Studien von W. Jay Dowling und anderen besprochen, die die Bedeutung der Kontur einer Melodie für die Gedächtnisrepräsentation und die Wahrnehmung allgemein unterstreichen. Auch für die Beurteilung von melodischen Ähnlichkeiten wird in vielen Modellen auf den Parameter der melodischen Kontur zurückgegriffen. Dabei wird die Kontur als Approximation an den Intervallverlauf einer Melodie konzeptualisiert, eine ungefähre Repräsentation, die der tatsächlichen Wahrnehmung der meisten Menschen weitgehend entspricht. Ähnlichkeitsmodelle, die von einem Vergleich der melodischen Konturen ausgehen, können deshalb als aussichtreiche Kandidaten für eine Übereinstimmung mit menschlichen Urteilen gelten. Als zwei Modelle, die sehr unterschiedliche Verfahren zur numerischen Kodierung der melodischen Kontur verwenden, werden Wolfram Steinbecks Polygonzug und Mark Schmucklers Fourrierreihen operationalisiert.

Steinbecks Polygonzug:

Stellte die Treppenkurve ein Modell zur Repräsentation von Tonhöhen und Dauern im kartesischen Koordinatensystem dar, so ist der Polygonzug eine Verallgemeinerung dieses Modells. Die anschaulichste Form des Polygonzuges ist eine geometrische Darstellung: Der Polygonzug besteht aus den einander anschließenden Geraden, die die Konturumkehrpunkte einer Melodie miteinander verbinden. Dieses Prinzip lässt sich an folgender Grafik aus Steinbecks Buch veranschaulichen, bei der im Gegensatz zur identischen Abbildung 4-2 die durchgezogenen Geraden von Interesse sind:

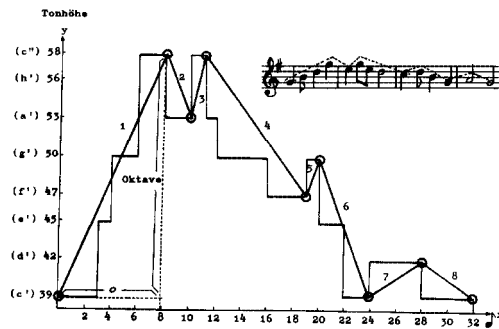


Abb. 4-4: Der Polygonzug einer melodischen Kontur, nach Steinbeck (1982), Abb. 47, S. 111.

Die Endpunkte der Geraden des Polygonzuges sind der erste Dauernwert des ersten Tones und die jeweils letzten Dauernwerte der Töne, an denen eine Konturumkehr stattfindet. Als Töne der Konturumkehr sind diejenigen definiert, deren numerische Tonhöhenwerte größer bzw. kleiner als die des vorangegangenen Tones und des nachfolgenden Tones sind. Die Behandlung von einfachen Wechselnoten wird weiter unten diskutiert. Es lässt sich die Steigung der Geraden zwischen den Konturumkehrpunkten berechnen nach:

$$b = \frac{TW_2 - TW_1}{D}$$

Die Differenz der aufeinanderfolgenden Tonhöhenwerte (TW) an den Konturumkehrpunkten wird durch die Anzahl der Dauernwerte (D) geteilt. Die vollständige Gleichung der Geraden ergibt durch Addition des Wertes der Ausgangstonhöhe der Geraden a :

$$y = a + bx$$

Gemäß dieser Formel wird für jede Dauerseinheit x einer Geraden ein zugehöriger Tonhöhenwert y errechnet. Die Reihe aller Tonhöhenwerte y aller Geraden ist die numerische Darstellung des Polygonzuges.

Da nach Steinbeck der Polygonzug vor allem dazu dienen soll, den Tonhöhenverlauf einer Melodie ‚geglättet‘, d.h. vereinfacht darzustellen, sollen nicht nur die Größe der einzelnen Intervallschritte nivelliert werden. Auch gegenläufige Wechselnoten, die den globalen Verlauf einer Melodie in der Wahrnehmung nur marginal beeinflussen, sollen, so Steinbeck, unberücksichtigt bleiben. Für die gegenläufigen Wechselnoten gibt Steinbeck eine erweiterte Definition:

Gegenläufige oder gegengerichtete Wechselnoten sind hier – in Erweiterung der üblichen Bedeutung des Terminus – jene Stellen, an denen die melodischen Richtungen einer Tonfolge durch einzelne in entgegengesetzte Richtung schreitende oder springende Tonhöhe (die bei Repetitionen auch mehrfach angeschlagen sein kann) unterbrochen und anschließend fortgesetzt werden. Melodische Richtungswechsel sind also erst gesichert, wenn mindestens zwei aufeinanderfolgende Intervalle gleichgerichtet sind. (Steinbeck, 1982, S. 113-114)

Da dies eine relativ strenge Definition von Wechselnoten ist, soll gleichzeitig noch eine weitere Variante des Wechselnotenausschlusses bei der Konstruktion des Polygonzuges verwendet werden, die aus der Menge der Richtungswechsel die Wechselnoten auf anderen Weise ausschließt und bei Müllensiefen und Frieler (2004) beschrieben ist. Diese Variante berücksichtigt einen Richtungswechsel als Konturumkehrpunkt nur dann nicht, wenn die Tonhöhenwerte vor und nach dem Richtungswechsel identisch sind. Dies entspricht der geläufigen Definition von Wechselnoten.

Zur Veranschaulichung für die verschiedenartige Behandlung von Wechselnoten durch die beiden Verfahren mögen die drei folgenden Beispiele von Tonhöhenverläufen in MIDI-Nummerierung gelten:

64, 65, 67, 65, 66

Der Ton 67 ist eine Wechselnote nach Steinbecks Definition und nach der von Müllensiefen und Frieler. Er käme nicht als Konturumkehrpunkt für die Berechnung des Polygonzuges in Frage.

64, 65, 67, 65, 64

Der Ton 67 ist eine Wechselnote im Sinne von Müllensiefen und Frieler, aber nicht in Steinbecks Sinne.

64, 65, 67, 66, 68

Der Ton 67 ist eine Wechselnote für Steinbecks Modell, aber ein echter Konturumkehrpunkt für das Müllensiefen-und-Frieler-Modell.

Mit Hilfe des Polygonzugmodells lässt sich also aus der Tonhöhenkontur einer Melodie auf mindestens zwei plausible Weisen ein Reihe numerischer Werte berechnen, die als Ersatztonreihen gegenüber den oben beschriebenen Werten des exakten Tonhöhenverlaufs bezeichnet werden können. In diesen Ersatztonreihen werden nur die deutlich wahrnehmbaren Verlaufsrichtungen berücksichtigt und einzelne, kleinere Unterschiede bei der Größe der Intervallschritte und den Wechselnoten ignoriert.

Wie bei der Treppenkurve des Tonhöhenverlaufs ist es auch hier sinnvoll, die Melodietöne vor der Berechnung der Werte der Geraden mit ihrer Tondauer zu gewichten. Denn diese Gewichtung hat, wie oben beschrieben, mit aller Wahrscheinlichkeit eine Entsprechung in der Wahrnehmung der Melodie und stellt gleichzeitig einen direkten Einfluß auf die Steigung der Geraden und damit die numerischen Werte dar. Es sollen deshalb nur Konturwerte miteinander verglichen werden, die an den Tondauern gewichtet wurden.

Als Ähnlichkeitsmaß schlägt Steinbeck wieder die durchschnittliche Differenz und den Produkt-Moment-Korrelationskoeffizienten r vor. Wie beim Vergleich der Treppenkurve zweier Tonhöhenverläufe ist es auch beim Vergleich der Polygonzüge geboten, die beiden Wertereihen gegeneinander zu verschieben, für jede Verschiebung einen Koeffizientenwert zu berechnen und den größten Koeffizientenwert als Wert für den wahren Ähnlichkeitszusammenhang zu nehmen (s.o.), wie es Steinbeck selbst vorschlägt (S.119-121).

Für das oben angeführte Beispiel der Testmelodie und ihrer Variante 7.4 ergeben sich folgende Ähnlichkeitswerte:

$$conspcst = 1$$

(Kontur Steinbeck, Pearson-Korrelation, auf Wertebereich 0-1 skaliert)

$$conpcst = 0,394$$

(Kontur Müllensiefen & Frieler, Pearson-Korrelation, auf Wertebereich 0-1 skaliert):

Der große Unterschied zwischen den beiden Ähnlichkeitswerten zeigt, wie entscheidend die Bestimmung und der Ausschluss von Wechselnoten sein kann. Nach der Steinbeck-schen Methode sind die Konturen der beiden Melodien bis auf die Wechselnoten identisch, nach Müllensiefen und Frieler bestehen deutliche Unterschiede.

Melodische Kontur als Fourierreihe nach Mark Schmuckler:

Ziel der ausführlichen Studie von Mark Schmuckler (1999) war die Auswahl eines Algorithmus zur Bestimmung der Ähnlichkeiten zwischen melodischen Konturen, der der menschlichen Wahrnehmung von Konturähnlichkeit am nächsten kommt. Dazu bestimmte Schmuckler die Ähnlichkeiten der Konturen von 20 verschiedenen Zwölftonreihen mit Hilfe von vier unterschiedlichen Modellen bzw. Algorithmen und verglich die Ähnlichkeitskoeffizienten mit den Bewertungen von 16 musikalisch gut vorgebildeten Versuchspersonen. Insofern ähnelt Schmucklers Studie stark dem Design der Ähnlichkeitsuntersuchung in dieser Arbeit. Wobei in dieser Arbeit die Ähnlichkeitsbestimmung nicht auf den Aspekt der Kontur beschränkt bleibt.

Als Verfahren, welches die größte Übereinstimmung mit den menschlichen Urteilen zeigte, ermittelte Schmuckler einen von ihm selbst entworfenen Algorithmus, bei dem die melodischen Konturen als komplexe Wellenformen behandelt werden, die durch die Fourieranalyse als eine Reihe numerischer Werte repräsentiert werden können. Auf Schmucklers

Untersuchung selbst, die eine ganze Reihe von diskussionswerten Punkten enthält (z.B. die Gleichsetzung von Komplexität und Konturähnlichkeit im Experiment), soll hier nicht näher eingegangen werden.

Verbindet man die einzelnen Töne einer einstimmigen Melodie im Notenbild mit einer durchgezogenen Linie, so erhält man ein wellenförmiges Gebilde, welches durch die Melodietöne (Punkte) auf der Wellenlinie und besonders an deren Wende- und Umkehrpunkten bestimmt wird. Mit Hilfe der Fourieranalyse ist es nun möglich, diese komplexe Welle in eine Reihe von überlagerten sinusförmigen Wellen zu zerlegen, die jeweils durch einen Amplituden- und einen Phasenwert gekennzeichnet werden können. Wie Schmuckler schreibt, stellen die Relationen zwischen den Amplituden und Phasen der Sinuskomponenten ein quantitatives Maß für die zyklischen Muster eines Signals (hier einer Melodiekontur) dar:

Considering all these frequency components, Fourier analysis provides a general description of the shape of the contour, taking into account both slow moving, low-frequency movement such as general trend, as well as any high frequency, point-to-point fluctuation in the contour. (Schmuckler, 1999, S. 302)

Schmuckler verwendet als numerische Darstellung der Tonhöhen nur deren Rangplätze gezählt von der tiefsten vorkommenden Tonhöhe aus. So erhält die tiefste Note den Wert 0, die zweittiefste den Werte 1 und die höchste von beispielsweise sieben verschiedenen Tonhöhen, die in einer Melodie vorkommen, den Wert 6. Diese Rangplatzvergabe scheint in Schmucklers Untersuchung sinnvoll, da er mit zwölfstönigen Melodien arbeitet, die sehr große Intervallsprünge beinhalten. Für die Konturen der Testmelodie und ihrer Variante 7.4 erhält man gemäß dieses Vorgehens folgende Wertereihen der Rangplätze:

Testmelodie: 5 4 3 4 3 2 2 5 5 3 1 1 5 4 3 4 3 2 5 1 0 7 7 5 5 6 6 5 4 5

Variante 7.4: 4 3 2 1 2 2 4 2 0 4 3 2 0 0 1 4 2 0 0 7 6 6 5 5 4 4 2 3

Unterwirft man diese Wertereihen jeweils einer diskreten Fourieranalyse, so erhält man $n/2$ Werte für die Amplituden und Phasen der beiden Konturen (hier also 15 bzw. 14 Werte).

Genau wie Steinbeck benutzt Schmuckler den Produkt-Moment-Korrelationskoeffizienten r und die durchschnittliche Differenz, um die Gleichförmigkeit bzw. den Unterschied zwischen den beiden Wertereihen zu bestimmen.

In der multiplen Regression, die Schmuckler schließlich durchführt (1999, S. 315-316), erweisen sich nach der Fourieranalyse nur die Werte des Amplitudenspektrums, aber nicht die Phasen als signifikanter Prädiktor. Aus diesem Grund sollen in dieser Arbeit auch nur

die Werte der Amplituden der Melodiekonturen für den Ähnlichkeitsvergleich herangezogen werden, für die aber jeweils beide Maßzahlen – der Korrelationskoeffizient und die durchschnittliche Differenz – verwendet werden. Die durchschnittliche Differenz wird in der Anwendung dieser Arbeit wie die oben definierte Differenz der Intervallabstände der Tonhöhenwerte *diff* auch an der Anzahl *n* der Wertepaare (hier: Amplitudenwerte) und dem Maximalwert aufeinanderfolgender Amplitudenwerte jeder Wertereihe ($\max(|y_{i+1} - y_i| + |x_{i+1} - x_i|)$) normiert. Das resultierende Maß wird mit *fourri* (das zweite ‚r‘ steht für die Rangkodierung der Tonhöhen, ‚i‘ steht für Intervalle, d.h. Differenzen) bezeichnet. Durch die Normierung ist der Wertebereich auf 0 bis 1 beschränkt. Genau wie bei der obigen Verwendung des Korrelationskoeffizienten wird auch hier bei der Korrelation die kürzere entlang der längeren Melodie verschoben und der maximale Ähnlichkeitswert aus diesen Korrelationen gleichlanger Wertereien als ‚wahrer‘ Ähnlichkeitswert interpretiert.

Schmuckler verzichtet in der eigentlichen Untersuchung zwar auf die Kodierung von Rhythmus bzw. Tondauern, schlägt in der abschließenden Diskussion seiner Ergebnisse aber eine Gewichtung der Tonhöhenrangplätze nach deren Dauern im Verhältnis zur kleinsten Dauerneinheit vor – also die gleiche Dauerngewichtung, die auch Steinbeck verwendet und die oben beschrieben wurde. Schmucklers Ähnlichkeitsmaße sollen in dieser Arbeit sowohl für die ungewichtete wie für die gewichtete Darstellung berechnet werden.

Für das Beispiel der beiden Melodien aus dem Versuch von Sloboda und Parker ergeben sich folgende Ähnlichkeitswerte:

Fourr = 0,944 (einfache Korrelation, da positiv = *fourrst*)

Fourrw = 0,92 (Korrelation auf gewichteten Ränge, da positiv = *fourrwst*)

Fourri = 0,79 (Differenz der Amplitudenwerte)

Die betrachteten Ähnlichkeitsmaße für den Vergleich von Melodiekonturen sind:

conspc Korrelationskoeffizient nach Pearson der gewichteten Konturersatzreihe, konstruiert nach Steinbeck

conspcst Korrelationskoeffizient nach Pearson der gewichteten Konturersatzreihe, konstruiert nach Steinbeck, auf den Wertebereich 0-1 transformiert

<i>conpc</i>	Korrelationskoeffizient nach Pearson der gewichteten Konturersatzreihe, konstruiert nach Müllensiefen und Frieler
<i>conpcst</i>	Korrelationskoeffizient nach Pearson der gewichteten Konturersatzreihe, konstruiert nach Müllensiefen und Frieler, auf den Wertebereich 0-1 transformiert
<i>fourr</i>	Korrelationskoeffizient nach Pearson der Amplituden der fouriertransformierten Tonhöhenränge
<i>fourrst</i>	Korrelationskoeffizient nach Pearson der Amplituden der fouriertransformierten Tonhöhenränge, auf den Wertebereich 0-1 transformiert
<i>fourrw</i>	Korrelationskoeffizient nach Pearson der Amplituden der fouriertransformierten, gewichteten Tonhöhenränge
<i>fourrwst</i>	Korrelationskoeffizient nach Pearson der Amplituden der fouriertransformierten, gewichteten Tonhöhenränge, auf den Wertebereich 0-1 transformiert
<i>fourri</i>	Durchschnittliche Differenz der Amplituden der fouriertransformierten Tonhöhenränge

C) Melodien als Strings von Symbolen: Die Edit Distance:

Der Algorithmus der sogenannten Edit Distance wurde 1965 vom russischen Wissenschaftler Vladimir Levenstein entworfen. Mit Hilfe des Algorithmus wird die minimale Anzahl an Transformationen bestimmt, die notwendig ist, um eine Zeichenkette in eine zweite zu transformieren. Die Edit Distance ist deshalb konzeptionell ein Unähnlichkeitsmaß. Als Transformationsschritte sind nur Löschungen, Einfügungen und Ersetzungen erlaubt.

Klassische Anwendungsgebiete für die Edit Distance, oder des mathematisch äquivalenten Konzeptes des String Alignments, sind der Vergleich von Sequenzen in der Molekularbiologie, die Ähnlichkeit von Buchstabenreihen (Sätzen oder Texten) in Textsuchsystemen und die Spracherkennung. Insofern ist es nicht verwunderlich, dass sich die meisten lehrbuchartigen Beschreibungen der Edit Distance in den Bereichen der Bioinformatik und der Computerwissenschaften finden (z.B. Gribskov & Devereux, 1991; Stephen, 1994; Gusfield, 1997, Lopestri & Wilfong, 1999). In der jüngeren Vergangenheit sind auch in dem in den letzten Jahren schnell wachsenden Sektor des sogenannten Music Information

Retrieval viele Modelle und Anwendungen entstanden, die auf eine Variante des Edit Distance Algorithmus zurückgreifen (z.B. Mongeau & Sankoff, 1990; McNab et al, 1996; Uitdenbogerd, 2002).

Wie der Edit-Distance-Algorithmus arbeitet, lässt sich am leichtesten an dem Vergleich von zwei Buchstabenstrings (Wörtern) erkennen. Das nachfolgende Beispiel ist an Gusfield (1997, S. 215ff) angelehnt. Dort werden auch die entsprechenden Lemmata, Beweise und Theoreme vorgeführt und ausführlich besprochen:

Aufgabe sei es, das Wort (d.h. den Buchstabenstring) *Schubert* in das Wort *Schoenberg* zu überführen. *I* steht im Beispiel für das Einfügen eines Buchstabens in den ersten String (insertion), *L* für das Löschen, und *E* bedeutet das Ersetzen eines Buchstabens im ersten String durch einen Buchstaben des zweiten Strings. *M* bezeichnet die Übereinstimmung zweier Buchstaben (match) und wird nicht als Operation gezählt.

Das sogenannte Edit Transcript, d.h. die Liste der minimal notwendigen Operationen, um SCHUBERT in SCHOENBERG zu verwandeln, ist:

Edit Transcript: M M ML L E M M M E
 S C H O E N B E R G
 S C H _ _ U B E R T

Es waren 2 Löschungen und 2 Ersetzungen notwendig, um den ersten Buchstabenstring in den zweiten zu überführen. Die Edit Distance wäre also 4, wenn man Löschungen und Ersetzungen gleich gewichtet.

Ein optimales Edit Transkript enthält die minimale Anzahl der notwendigen Operationen. Mit Hilfe eines einfachen rekursiven Algorithmus ist das optimale Edit Transkript ermittelbar. Der Algorithmus, der die kleinste Distance $D(i, j)$ zweier Strings (S_1 und S_2) bestimmt, geht von zwei Voraussetzungen aus:

$D(i, 0) = i \Rightarrow$ die Distanz eines Strings der Länge i und eines zweiten der Länge 0 ist gleich i (i Löschungen).

$D(0, j) = j \Rightarrow$ die Distanz eines Strings der Länge 0 und eines zweiten der Länge j ist gleich j (j Einfügungen).

Die rekurrente Relation der Distanzen zweier Strings der Länge i und j ist die folgende:

$$D(i, j) = \min [D(i-1, j) + 1, D(i, j-1) + 1, D(i-1, j-1) + t(i, j)]$$

Dabei hat $t(i, j)$ den Wert 1, wenn $S_1(i)$ ungleich $S_2(j)$ und den Wert 0, wenn $S_1(i) = S_2(j)$ ist.

Allerdings steigt die Anzahl der rekursiven Operationen exponentiell mit der Länge der beiden Strings und wird so für längere zu vergleichende Strings ineffektiv (s. Gusfield, 1997, S.219), wenn man Rekursion als gewöhnliche ‚Top-Down-Programmierung‘ benutzt.

So wird auf die Methode der Tabellenberechnung – einen Ansatz der dynamischen Programmierung – zurückgegriffen, bei der der Lösungsraum mit einem Traceback-Verfahren vollständig durchsucht und so die optimale Lösung, d.h. die kleinste Edit Distance, ermittelt wird. Eine Tabelle für die Distanz zwischen SCHUBERT und SCHOENBERG wird nach dem folgenden Schema gefüllt:

Die oberste Zeile und die Spalte ganz links (innerhalb der Umrahmung) füllen sich aus den beiden oben genannten Ausgangsvoraussetzungen.

Die übrigen Zellen werden zeilenweise von oben nach unten und innerhalb der Zeilen von links nach rechts gefüllt. Die Vorschrift für das Füllen einer Zelle besteht darin, den niedrigsten der drei möglichen Zellenwerte zu wählen, die nach folgenden Vorschriften ermittelt werden (s. formale Definition oben):

- gehe in der selben Zeile eine Zelle nach links und addiere 1,
- gehe in der selben Spalte eine Zelle nach oben und addiere 1,
- gehe zur Zelle diagonal links oben und addiere 1, wenn in der Ausgangszelle der Spalten- und Zeilenbuchstaben nicht übereinstimmen oder addiere 0, wenn in der Ausgangszelle der Spalten- und Zeilenbuchstabe übereinstimmen.

D(i, j)			S	C	H	U	B	E	R	T
		0	1	2	3	4	5	6	7	8
	0	0	1	2	3	4	5	6	7	8
S	1	1	↖ 0	1	2	3	4	5	6	7
C	2	2	1	↖ 0	1	2	3	4	5	6
H	3	3	2	1	↖ 0	1	2	3	4	5
O	4	4	3	2	↑ 1	↖ 1	2	3	4	5
E	5	5	4	3	↑ 2	↖ ↑ 2	2	2	3	4
N	6	6	5	4	3	↖ ↑ 3	3	3	3	4
B	7	7	6	5	4	4	↖ 3	4	4	4
E	8	8	7	6	5	5	4	↖ 3	4	5
R	9	9	8	7	6	6	5	4	↖ 3	4
G	10	10	9	8	7	7	6	5	4	↖ 4

Tabelle 4-1: Tabelle zur Berechnung der Edit Distance zwischen den Buchstabenstrings SCHUBERT und SCHOENBERG

Die Frage nach der minimalen Anzahl der Transformationsoperationen kann mit dieser Tabelle durch die Frage nach dem optimalen Weg von Zelle (T G) zu Zelle (0 0) ersetzt werden (Traceback-Verfahren). Dazu werden Zeiger in den einzelnen Zellen nach folgenden Regeln gesetzt:

Setze einen Zeiger von Zelle (i, j) zu Zelle $(i - 1, j)$, wenn $D(i, j) = D(i - 1, j) + 1$

Setze einen Zeiger von Zelle (i, j) zu Zelle $(i, j - 1)$, wenn $D(i, j) = D(i, j - 1) + 1$

Setze einen Zeiger von Zelle (i, j) zu Zelle $(i - 1, j - 1)$, wenn $D(i, j) = D(i - 1, j - 1) + t(i, j)$

Folgt man allen möglichen Pfaden von Zelle (T G) zu Zelle (0 0), so erhält man alle optimalen Alignments bzw. Edit Transkripts. In diesem Fall sind das die Lösungen:

Edit Transkript: M M M L L E M M M E
 S C H O E N B E R G
 S C H _ _ U B E R T

Edit Transkript: M M M E L L M M M E

S C H O E N B E R G

S C H U _ _ B E R T

Edit Transkript: M M M L E L M M M E

S C H O E N B E R G

S C H _ U _ B E R T

In jedem Fall sind zwei Löschungen und zwei Ersetzungen notwendig, d.h. die Edit Distance der beiden Buchstabenstrings ist 4. Dieser Wert findet sich in Zelle (T, G) , 'unten rechts' in der Tabelle.

Analog lässt sich die Berechnung der Edit Distance für musikalische Symbole, z.B. Tonhöhen, Intervalle oder Tondauern errechnen. Ein Beispiel dafür wären die Tonhöhen der ersten beiden Takte der Testmelodie und der Variante 7.4, für die sich der folgende Lösungsraum ergibt:

D(i, j)	Original		a	g	f	g	f	e	e	a	a	f	d	d
Variante7.4		0	1	2	3	4	5	6	7	8	9	10	11	12
	0	0	1	2	3	4	5	6	7	8	9	10	11	12
a	1	1	↖0	1	2	3	4	5	6	7	8	9	10	11
g	2	2	1	↖0	1	2	3	4	5	6	7	8	9	10
f	3	3	2	1	↖0	1	2	3	4	5	6	7	8	9
e	4	4	3	2	1	↖1	2	2	3	4	5	6	7	8
f	5	5	4	3	2	2	↖1	←2	3	4	5	5	6	7
f	6	6	5	4	3	3	2	↖2	←↖3	4	5	5	6	7
a	7	7	6	5	4	4	3	3	↖3	←↖3	←4	5	6	7
f	8	8	7	6	5	5	4	4	4	4	↖4	←↖4	←5	←6

Tabelle 4-2: Tabelle zur Berechnung der Edit Distance zwischen den Anfangsphrasen der Testmelodie von Sloboda und Parker (1985) und der Variante 7.4.

Die Edit Distance der zwei ersten Takte aus beiden Melodien ist also 6. Um den Ähnlichkeitswert der Edit Distance auf den Bereich der übrigen Ähnlichkeitsmaße zwischen 0 und 1 zu beziehen, wird die Zahl der notwendigen Transformationen noch durch die Zahl der maximal möglichen Transformationen, d.h. die Zahl der Elemente der längeren der beiden

Symbolstrings geteilt. Dieser Wert wird von 1 subtrahiert, sodass folgender Ähnlichkeitswert für das Beispiel zustande kommt:

$$1 - 6/12 = 0,5$$

Der vorgeführte Algorithmus wird in der Literatur häufig als *Global Alignment* bezeichnet und ist v.a. für den Vergleich von zwei Strings von ungefähr gleicher Länge geeignet. Er stellt allerdings nur die einfachste Variante von (Un-)Ähnlichkeitsberechnungen mit Hilfe des Alignments von zwei Symbolstrings dar. Als Erweiterungen bzw. Varianten besteht z.B. die Möglichkeit, Löschungen und Einfügungen anders zu bewerten (z.B. mit -1) als Ersetzungen (z.B. -2) und Übereinstimmungen (z.B. 3). Darüber hinaus ist es möglich, Alphabete der Buchstabenstrings zu definieren und die Ersetzung bestimmter Symbole durch andere bestimmte Symbole besonders zu gewichten. Im vorliegenden Beispiel könnte die Ersetzung von dunklen Vokalen durch dunkle Vokale (z.B. O durch U) geringer gewichtet werden. Im Bereich der Musik könnte das Ersetzen einer großen Terz durch eine kleine und umgekehrt geringer gewichtet werden oder das Ersetzen von melodischen Stufen und Sprüngen jeweils gegeneinander.

In vielen Anwendungen müssen Strings von deutlich unterschiedlicher Länge miteinander verglichen werden, z.B. bei der Suche nach einzelnen Motiven in der Melodiedatenbank mit vollständigen Melodien und Themen. Oder die Aufgabe besteht darin, Substrings mit hoher Ähnlichkeit aus zwei sehr langen Ausgangsstrings zu finden. Als Beispiele sind denkbar das Auffinden gleicher Abschnitte in DNA-Sequenzen oder der Vergleich zweier Gitarrensoli die teilweise ähnliche Licks verwenden. Die Techniken zur Lösung dieser Probleme bespricht Gusfield unter der Bezeichnung *Local Alignment* (1997, S. 230ff). An dieser Stelle soll hierauf jedoch nicht weiter eingegangen werden, da es sich bei den als Versuchsmaterial verwendeten Melodien um in etwa gleich lange Abschnitte handelt und das oben beschriebene globale Alignment als Lösungsansatz ausreicht.

Bei der Implementierung für den vorliegenden Versuch wurde der beschriebene Algorithmus des *Global Alignment* implementiert. Verglichen werden sowohl die Reihe der absoluten (transponierten) Tonhöhen, die Folge der Intervalle sowie der quantisierten Tondauern (s.u.) und der harmonische implizite Inhalt der einzelnen Takte (s.u.).

Die verwendeten Ähnlichkeitsmaße sind:

rawed Edit Distance der (transponierten) Tonhöhenwerte

rawedw Edit Distance der (transponierten) gewichteten Tonhöhenwerte

<i>coned</i>	Edit Distance der gewichteten Konturersatzreihen, konstruiert nach Müllensiefen und Frieler
<i>consed</i>	Edit Distance der gewichteten Konturersatzreihen, konstruiert nach Steinbeck
<i>diffed</i>	Edit Distance der Intervallfolgen
<i>rhythfuzz</i>	Edit Distance der quantisierten (fuzzyfizierten) Dauernfolgen (s.u.)
<i>harmcore</i>	Edit Distance der impliziten Harmonien der einzelnen Takte (s.u.)

D) Melodien als Abfolge von Tondauern:

Der Parameter *Rhythmus* ist zwar schon durch die Gewichtung der Tonhöhen an den zugehörigen Dauern in den bisherigen Modellen der Tonhöhenverläufe, der Kontur und der Symbolstrings implizit berücksichtigt worden, dort aber immer mit dem Tonhöhenparameter verbunden gewesen. Mit den bisher vorgestellten Modellen ist es deshalb nicht möglich, einen bedeutsamen Ähnlichkeitswert für zwei Melodien zu erhalten, die im Tonhöhenverlauf verschieden, im Rhythmus aber identisch sind. Der Rhythmus kann jedoch in vielen Situationen der entscheidende Parameter sein, der melodische Ähnlichkeit beim Hörer vermittelt. Als Beispiele wären zu nennen die motivisch-thematische Verarbeitung in Werken der westlichen Kunstmusik, bei der abschnittsweise Intervallstruktur und Kontur stark verändert sein können und die motivische Kohärenz vor allem über einen gleichbleibenden Rhythmus vermittelt wird. Ein weiteres Beispiel sind die gesungenen Wiedergaben von Melodien durch sogenannte ‚jugendliche Brummer‘, „die ein ganzes Lied auf einem Ton brummen“ (Bruhn, 1991, S. 288-289) oder eine Kontur singen, „die aber der vorgegebenen Liedkontur überhaupt nicht [entspricht]“ (S. 289). Auch hier wäre der Rhythmus der einzig verwertbare Parameter.

Da in der Literatur exklusiv rhythmische Ähnlichkeitsmaße sehr selten sind, sollen zwei Maße kurz vorgestellt werden, bei deren Konstruktion eine Orientierung an bestehenden Ähnlichkeitsmaßen für andere Parameter stattgefunden hat.

Abfolge von Dauernklassen:

Das erste Maß, der Vergleich der Abfolge von Dauernklassen, beruht auf der Annahme, dass musikalisch nicht vorgebildete Personen musikalische Dauern in wenigen Dauernklassen kategorisch wahrnehmen. Drake und Bertrand (Drake & Bertrand, 2001, S. 24-25) postulieren eine Dreiteilung in a) Dauern derselben Länge wie eine gegebene Refe-

renzdauer, b) halb so lange Dauern und c) doppelt so lange Dauern. Sie führen zur Stützung ihrer These musikpsychologische Literatur an (z.B. Clarke, 1987; Schulze, 1989; Parncutt, 1994) und argumentieren sogar, dass diese Art der kategorischen Dauernwahrnehmung den Status einer musikalischen Universalie haben könne.

Entsprechend werden für das vorliegende Ähnlichkeitsmaß die Dauern in fünf Klassen eingeteilt. Zusätzlich zu den drei genannten kommen noch die Klassen „< halbe Dauer“ und „> doppelte Dauer“ hinzu, um das ganze Spektrum an möglichen Dauern und Pausen abdecken zu können.

Als Referenzdauer wird die häufigste Dauer einer Melodie definiert (Modus der Dauern). In dem Fall, dass eine Melodie mehrere Dauernmodi aufweist bzw. die beiden zu vergleichenden Melodien verschiedene Modi haben, wird auf der Basis jeden Dauernmodus' ein eigener Dauernvergleich gerechnet. Von diesen verschiedenen Dauernvergleichen wird der maximale Ähnlichkeitswert als endgültiger Ähnlichkeitswert definiert.

Für die Klassengrenzen der Dauernkategorien nennen Drake und Bertrand leider keine numerischen Werte. Sinnvoll erscheint es aber in jedem Fall, eine Schwankungsbreite der Dauern um die exakten Verhältnis 1:2, 1:1 und 2:1 anzunehmen. Aufgrund der musikalischen Erfahrung, dass Laien eher dazu tendieren, einen Beat zu verschleppen bzw. die einzelnen Schläge zu spät zu treffen, wurden Klassengrenzen nicht im gleichen Abstand von ihrem Klassenverhältniswert gelegt, sondern etwas auf der Zeitachse nach ‚rechts‘, d.h. zu den längeren Dauern hin verschoben. Die Grenzen der Dauernklassen wurden also wie folgt festgelegt:

Dauernklasse	< 1:2	1:2	1:1	2:1	>2:1
Grenzen (in Vielfachen der Referenzdauer)	0 – 0,45	0,45 – 0,9	0,9 – 1,8	1,8 – 3,3	>3,3

Tabelle 4-3: Grenzen der Dauernklassen gemäß der eigenen Dauernklassifizierung.

Die Dauern der von Sloboda und Parker verwendeten Testmelodie und der Variante 7.4 werden also zur Abfolge folgender Dauernklassen. Pausen werden dabei zu den Dauern der Vorgängernote addiert:

Testmelodie (häufigste Dauernklasse: Achtel):

2 1 2 1 1 1 1 1 1 2 1 2 1 2 1 1 2 2 > 2 2 1 1 2 1 1 1 1 1 2

Variante 7.4 (häufigste Dauernklasse: Achtel):

2 1 2 1 1 > 2 2 2 > 2 2 1 2 1 1 2 2 > 2 1 2 1 1 1 1 1 2 2 1 1

Als Vergleichsmaß für zwei Dauernfolgen diene die oben schon besprochene Edit Distance, die auch hier auf den Wertebereich zwischen 0 und 1 normiert wurde. Für die beiden Dauernfolgen wurde die minimale Edit Distance berechnet, wofür 10 Transformationen notwendig waren. Dies ergibt einen Ähnlichkeitswert durch die Edit Distance von 0,67.

„Gaußifizierung“ von Toneinsatzzeitpunkten:

Das zweite Ähnlichkeitsmodell für den Rhythmusvergleich stammt von Frieler (Frieler, 2004, 2005) und beruht auf der Verwandlung der realen in einer Melodie vorkommenden Toneinsatzwerte in eine integrierbare Funktion, deren Werte sich für die Berechnung von Skalarprodukten und damit Korrelationen eignen. Frieler vergleicht also nicht direkt die Tondauern, sondern die Tonanfangswerte zweier Melodien miteinander, die natürlich implizit auch Längen-(=Dauern-)informationen enthalten. Auf die mathematische Herleitung der von ihm so genannten „Gaußifizierung“ von Toneinsatzpunkten und die Berechnung von Ähnlichkeitswerten zweier Onsetfolgen mittels der Kreuzkorrelation soll hier nicht weiter eingegangen werden. Die entsprechenden Details finden sich in der Originalveröffentlichung (Frieler, 2004).

Die durch die Gaußifizierung gewonnenen Ähnlichkeitswerte können sich zwischen 0 und 1 (=max. Ähnlichkeit) bewegen.

Genau wie bei der Berechnung des Korrelationskoeffizienten r , wird auch bei der Verwendung der Kreuzkorrelation die kürzere Melodie entlang der längeren verschoben, um den möglichen Effekt von Auftakten auszugleichen. Für jede Verschiebung wird ein eigener Wert berechnet und der maximale Wert aller Verschiebungen als der wahre Ähnlichkeitswert definiert.

Die Kreuzkorrelation der gaußifizierten Onsetwerte ergab für das Beispiel der Testmelodie und ihrer Variante 7.4 den Wert 0,64 (bei einer Verschiebung der Variante 7.4 um 2 Einheiten nach rechts).

Als reine Rhythmusmaße sollen somit die beiden folgenden dienen:

<i>rhythfuzz</i>	Edit Distance der zu fünf Klassen zusammengefassten Tondauern
<i>rhythgaus</i>	Kreuzkorrelation der ‚gaußifizierten‘ Onsetwerte

E) Melodien als Mengen von geordneten Tupeln melodischer Elemente (n-gram-Ansatz):

Der wahrscheinlich jüngste Ansatz zur Ähnlichkeitsbestimmung von Melodien ist der sogenannte *n-gram*-Ansatz. Ähnlich wie die Edit Distance wurde der *n-gram*-Ansatz für Such- und Vergleichsprozesse in anderen Domänen, wie großen Textdatenbanken schon länger erfolgreich eingesetzt. Für den Bereich von *Melodie-Retrieval-Systemen* wurden die *n*-grams vermutlich jedoch erstmals von Downie 1999 bearbeitet und systematisch untersucht (Downie, 1999).

Ein *n-gram* stellt nach dieser Aufarbeitung für den Bereich der Melodien eine geordnete Reihe von Elementen dar, z.B. Tonhöhen, Intervalle, diatonische Intervallklassen, Tondauern, Bewegungsrichtungen oder Kombinationen aus diesen elementaren Dimensionen. Die Länge des *n-grams* ist dabei im Prinzip frei wählbar. Für den Vergleich von melodischen Daten hat sich jedoch eine *n-gram*-Länge von 3 bis 8 Elementen als am geeignetsten erwiesen (Uitdenbogerd & Zobel, 2001).

So lassen sich die jeweils ersten Phrasen der Testmelodie von Sloboda und Parker und deren Variante 7.4 folgendermaßen als Folge von Tonhöhen-*n*-grams der Länge 3, 4 und 5 darstellen:

1. Phrase Test- melodie	3-grams	4-grams	5-grams	1. Phrase Variante 7.4	3-grams	4-grams	5-grams
a	agf	agfg	agfgf		agf	agfe	agfef
g	gfg	gfgf	gfgfe	g	gfe	gfef	gfeff
f	fgf	fgfe	fgfee	f	fef	fefe	feffa
g	gfe	gfee	gfeea	e	eff	effa	effaf
f	fee	feea	feea	f	ffa	ffaf	
e	eea	eeaa	eeaaf	f	faf		
e	eaa	eaaf	eaafd	a			
a	aaf	aafd	aafdd	f			
a	afd	afdd					
f	fdd						
d							
d							

Tabelle 4-4: 3-grams der Anfangsphrasen der Testmelodie von Sloboda und Parker und der Variante 7.4.

Für den Ähnlichkeitsvergleich zweier Melodien stehen nun mindestens drei unterschiedliche Maße zur Verfügung, die alle – z.B. bei Uitdenbogerd (2002) – sehr gut beschrieben sind:

Summe gemeinsamer n-grams (Sum Common):

Hier werden die *n-grams* gesucht, die sowohl in Melodie 1 als auch Melodie 2 vorkommen. Die Häufigkeiten jedes dieser gemeinsamen *n-grams* pro Melodie werden aufsummiert und schließlich durch die maximale Anzahl aller *n-grams* beider Melodien geteilt. In der Formelschreibweise lässt sich dies wie folgt darstellen. Die Summe der Häufigkeiten c der gemeinsamen *n-grams* ist:

$$c(s, t) = \sum_{\tau \in S_n \cap t_n} f_s(\tau) + f_t(\tau)$$

Dabei steht τ für die gemeinsamen *n-grams* in beiden Melodien und $f_s(\tau)$ und $f_t(\tau)$ für deren absolute Häufigkeiten in den Melodien s bzw. t . Normiert wird diese Summe an der Anzahl aller möglichen *n-grams* der beiden Melodien. Die Anzahl der möglichen *n-grams* ist gegeben durch $|s| - n + 1$ für eine Melodie und durch $|s| + |t| - 2(n - 1)$ für beide Melodien.

Als Ähnlichkeitsmaß ergibt sich folglich:

$$\sigma(s, t) = \frac{c(s, t)}{|s| + |t| - 2(n - 1)}$$

Für die beiden Anfangsphrasen im Beispiel ergeben sich folgende gemeinsame *3-grams*: *agf*, *gfe*. Beide *3-grams* kommen in beiden Melodieanfängen je einmal vor. Bei 10 bzw. 6 verschiedenen *3-grams* der beiden Melodieanfänge ergibt sich somit ein Ähnlichkeitswert von 0,25.

Ukkonen-Maß:

Das Ukkonen-Maß zählt die Unterschiede in den absoluten Häufigkeiten aller *n-grams* in beiden Melodien. Es ist also von der Konzeption her ein Unterschiedsmaß und gegensätzlich zum *Sum Common*-Maß konstruiert. Um es auf den Wertebereich von 0 bis 1 zu skalieren, wird der an der maximalen Differenz normalisierte Ukkonen-Werte von 1 abgezogen. Als Formel lässt sich das hier verwendete Ukkonen-Maß folgendermaßen ausdrücken:

$$u(s, t) = \sum_{\tau \in S_n \cup t_n} |f_s(\tau) - f_t(\tau)|$$

Als Ähnlichkeitsmaß resultiert also:

$$\sigma(s, t) = 1 - \frac{u(s, t)}{|s| + |t| - 2(n - 1)}$$

Für das vorliegende Beispiel finden sich 12 *3-grams*, die nicht in beiden Melodien und je einmal vorkommen (Differenz je 1) sowie 2 *3-grams*, die in beiden Melodien einmal vorkommen (Differenz je 0). Insgesamt ist der Unterschied in den Häufigkeiten der *3-grams* 12. Da insgesamt 16 *3-grams* in beiden Melodien zusammen vorhanden sind, ergibt sich ein Ähnlichkeitswert von $1 - \frac{12}{16} = 0,25$.

Coordinate Matching:

Für das *Coordinate Matching*-Maß (auch *Count Distinct*-Maß genannt) wird die Anzahl der gemeinsamen *n-grams* in beiden Melodien gezählt, ohne deren Häufigkeiten zu beachten. Diese Anzahl wird durch die Anzahl der *n-grams* des längeren Strings geteilt:

$$\sigma(s, t) = \frac{\sum_{\tau \in s_n \cap t_n} 1}{\max(\#s_n, \#t_n)}$$

Für die Beispielmelodien lassen sich 2 gemeinsame *3-grams* finden. Der längere Melodieanfang enthält 10 *3-grams*. Als Ähnlichkeit ergibt sich demnach 0,2.

Als Ausgangsmaterial für die *n-gram*-Maße wurden vier verschiedene Datentypen verwendet: die Intervalle, die oben definierten Dauernklassen, Intervallrichtungen (Parsons Code) und fuzzyfizierte Intervallklassen.

Der sogenannte Parsons Code betrachtet nur die drei möglichen Richtungen (auf, ab, gleich) der Intervallbewegungen. Er bildet damit die Intervallkontur auf grobe Weise ab.

Ähnlich wie für die Tondauern schien es sinnvoll, auch für die Intervalle eine reduzierte (mid level) Repräsentation zu schaffen, die in ihrer Genauigkeit zwischen dem Parsons Code (low level) und den exakten Intervallen (high level) liegt. Hierfür wurden die 23 möglichen Intervalle (einer Oktave) auf 9 Kategorien reduziert: Die Zuordnung folgte dem in der Tabelle dargestellten Schema:

Intervallklasse	Zugeordnete Intervalle
Prim	Prim
Sekunde aufwärts	Kleine, große Sekunde aufwärts
Terz aufwärts	Kleine, große Terz aufwärts
Quart/Quint aufwärts	Quarte, Tritonus, Quinte aufwärts
Große Intervalle aufwärts	Kleine, große Sexte, kleine, große Septime, Oktave und größere Intervalle aufwärts
Sekunde abwärts	Kleine, große Sekunde abwärts
Terz abwärts	Kleine, große Terz abwärts
Quart/Quint abwärts	Quarte, Tritonus, Quinte abwärts
Große Intervalle abwärts	Kleine, große Sexte, kleine, große Septime, Oktave und größere Intervalle abwärts

Tabelle 4-5: Intervallklassen und zugehörige Intervalle für die auf Intervallklassen basierenden *n*-gram-Maße.

Mit Hilfe dieser Intervallklassen lassen sich melodische Motive als identisch erkennen, die sich nur durch diatonische Anpassungen auf einer anderen Tonstufe unterscheiden. So unterscheiden sich beispielsweise Dux und Comes barocker Fugen in ihrer Intervallabfolge in der Regel nur durch halbtönige Anpassungen, die der Versetzung des Comes auf die fünfte Tonstufe geschuldet sind. Auch melodische Phrasen, die z.B. im Verlauf der Durchführung einer Sonate motivisch-thematisch verarbeitet werden, weisen häufig noch dieselben diatonischen Intervallklassen auf.

Mit dem *n*-gram-Ansatz lässt sich das Auftreten von kurzen Folgen von Elementen in zwei Ketten von Elementen beschreiben. Musikalisch interpretiert können die kurzen Folgen für melodische oder rhythmische Motive stehen, die in zwei Melodien vorkommen können. Gerade das Vorkommen von gemeinsamen Motiven kann bei einem ansonsten stark veränderten melodischen Kontext für die menschliche Ähnlichkeitswahrnehmung verantwortlich sein.

Für die Ähnlichkeitsbestimmungen im Rahmen dieser Arbeit wurden die folgenden auf dem *n*-gram-Konzept basierenden Ähnlichkeitsmaße verwendet, wobei jedes Maß für *n*-grams der Längen 3,4,5,6,7, und 8 getrennt berechnet wurde. Der maximale Ähnlichkeitswert wurde als ‚wahre‘ Ähnlichkeit für die beiden zu vergleichenden Melodien interpretiert. Die verwendeten *n*-gram-Maße sind:

<i>ngrsumco</i>	Sum Common-Maß auf Basis der Intervalle
<i>ngrukkon</i>	Ukkonen-Maß auf Basis der Intervalle
<i>ngrcoord</i>	Coordinate matching (count distinct)-Maß auf Basis der Intervalle

<i>ngrsumcr</i>	Sum Common-Maß auf Basis der Intervallrichtungen (Parsons Code)
<i>ngrukkor</i>	Ukkonen-Maß auf Basis der Intervallrichtungen (Parsons Code)
<i>ngrcoorr</i>	Coordinate matching (count distinct)-Maß auf Basis der Intervallrichtungen (Parsons Code)
<i>ngrsumcf</i>	Sum Common-Maß auf Basis der Intervallklassen
<i>ngrukkof</i>	Ukkonen-Maß auf Basis der Intervallklassen
<i>ngrcoorf</i>	Coordinate matching (count distinct)-Maß auf Basis der Intervallklassen
<i>ngrsumfr</i>	Sum Common-Maß auf Basis der Dauernklassen
<i>ngrukkfr</i>	Ukkonen-Maß auf Basis der Dauernklassen
<i>ngrcoofr</i>	Coordinate matching (count distinct)-Maß auf Basis der Dauernklassen

F) Melodien als Abfolge von Tonalitätswerten:

Tonalität ist bei der Wahrnehmung, Erinnerung und auch beim Ähnlichkeitsvergleich von ‚echter‘ Musik, d.h. hier mehrstimmiger oder akkordisch begleiteter Musik, mit Sicherheit ein sehr wichtiger Faktor, der bei der algorithmischen Berechnung von Ähnlichkeiten auf jeden Fall berücksichtigt werden sollte. Im vorliegenden Fall werden einstimmige Melodien miteinander verglichen. Die harmonische Information bzw. die impliziten Tonalitäten können somit nur aus der einstimmigen Linie induziert werden. Auch bei der Verarbeitung von einstimmigen Melodien weisen die Ergebnisse von Dowling (1978) und anderen darauf hin, dass die Wahrnehmung von Tonalität oder einer zugrundeliegende Skala, welche durch die einzelnen Melodietöne induziert wird, von großer Bedeutung ist. Diese implizite harmonische Information soll für die Modellierung der menschlichen Ähnlichkeitsurteile herangezogen werden.

Für die Bestimmung von harmonischer Information oder Tonalitätsinformation auch aus einstimmigen melodischer Linien sind mehrere Algorithmen entworfen worden (z.B. Longuet-Higgins & Steedman, 1971; Holtzmann, 1977; Krumhansl, 1990; Temperley, 1997; Krumhansl & Toiviainen, 2001). Für die Bestimmung der Tonalitäten der Melodien in dieser Arbeit soll nur einer dieser Algorithmen (Krumhansl, 1990) zur Anwendung kommen, der jedoch diverse Vorteile aufweist: Das im folgenden ‚Krumhansl-Algorithmus‘ genannte Verfahren ist relativ simpel aufgebaut, sodass es leicht zu implementieren ist und eine konzeptuelle Transparenz aufweist. Es beruht auf empirischen Daten, die von Hörern erhoben wurden und hat deshalb eine musikpsychologische Fundierung. Es ist an Musik verschiedener Stile getestet worden; und es liefert einen differenzierten Output in

Form eines 12-dimensionalen Tonalitätsvektors, was der vermuteten differenzierten Tonalitätswahrnehmung von geschulten Hörern entspricht.

Bei der Konstruktion des Tonalitätsalgorithmus geht Krumhansl von den Gewichten der 12 chromatischen Tonstufen in einer Dur- bzw. Molltonalität aus, die sie gemeinsam mit Kessler mit Hilfe der von ihr entwickelten Probe-Tone-Method ermittelt hatte (Krumhansl & Kessler, 1982). Auch wenn das Probe-Tone-Verfahren nicht unumstritten ist, so scheinen doch die Gewichte der einzelnen Tonstufen im Lichte allgemeiner musiktheoretischer Kenntnisse plausibel:

Tonstufe (chromatisch)	Gewichte (C-Dur)	Gewichte (c-moll)
C	6,35	6,33
Cis	2,23	2,68
D	3,48	3,52
Dis	2,33	5,38
E	4,38	2,60
F	4,09	3,53
Fis	2,52	2,54
G	5,19	4,75
Gis	2,39	3,98
A	3,66	2,69
Ais	2,29	3,34
H	2,88	3,17

Tabelle 4-6: Tonstufen der chromatischen Tonleiter in Tonalitätsgewichte in Dur und moll nach Krumhansl und Kessler am Beispiel C.

Per Definition sind diese Gewichte C-Dur und c-moll zugeordnet. Die Gewichte für die 11 anderen Dur- und Moll-Tonarten werden durch ‚Verschieben‘ der Reihenfolge der Gewichte erreicht. So hat der Vektor für Des-Dur die Werte: 2,88; 6,35; 2,23; 3,48; 2,33; 4,38; 4,09; 2,52; 5,19; 2,39; 3,66; 2,29; d.h. die Werte werden um eine Stelle nach rechts verschoben und an die erste Stelle rückt der letzte Wert des C-Dur-Vektors. Alle anderen Tonalitätsvektoren werden nach demselben Prinzip gebildet. Es ergeben sich für 12 Dur- und 12 Moll-Tonarten 24 Tonalitätsvektoren.

Mit diesen Gewichten werden nun für die Tonalitätsbestimmung die empirischen Tonhöhenwerte einer zu bestimmenden Melodie verglichen. Dazu werden die Tonhöhenwerte der Melodie an ihren zugehörigen Dauern gewichtet, ähnlich wie oben im Abschnitt zum Tonhöhenvergleich beschrieben. Es wird die Anzahl der 16tel gezählt, für die eine bestimmte Tonhöhe erklingt. Für die beiden ersten Takte der Testmelodie aus Sloboda und Parker und

deren Variante 7.4 ergeben sich folgende Werte (Pausen werden zur Vorgängernote hinzugezählt):

Tonhöhen (Teschmelodie)	Anzahl 16tel	Tonhöhen (Va- riante 7.4)	Anzahl 16tel
Takt 1			
C	0	C	0
Cis	0	Cis	0
D	0	D	0
Dis	0	Dis	0
E	0	E	2
F	6	F	6
Fis	0	Fis	0
G	4	G	2
Gis	0	Gis	0
A	6	A	6
Ais	0	Ais	0
H	0	H	0
Takt 2			
C	0	C	0
Cis	0	Cis	0
D	6	D	0
Dis	0	Dis	0
E	4	E	0
F	2	F	12
Fis	0	Fis	0
G	0	G	0
Gis	0	Gis	0
A	4	A	4
Ais	0	Ais	0
H	0	H	0

Tabelle 4-7: Tonhöhen und deren Dauer (in 16teln) der Testmelodie von Sloboda und Parker und deren Variante 7.4.

Zur Ermittlung der Tonalität der insgesamt vier Takte werden nun die Werte der 16tel, an denen die einzelnen Tönen klingen (Spalten 2 und 4), mit den insgesamt 24 Tonalitätsvektoren nach Pearson-Bravais korreliert. Die höchste Korrelation wird als Hinweis auf den ‚wahren‘ Tonalitätsvektor gewertet.

Im Beispiel der beiden Anfangstakte ergaben sich für die Takte 1 und 2 des Originals die höchsten signifikanten Korrelationen mit den Tonalitätsvektoren von F-Dur ($r=0,632$) bzw. d-moll ($r=0,789$). Für die Takte 1 und 2 der Variante 7.4 waren die höchsten Korrelationen in beiden Fällen die mit dem Vektor von F-Dur ($r=0,638$ bzw. $r=0,737$).

In der Implementierung für diese Arbeit wurde für jeden Takt einer Melodie die Tonalität ermittelt. Dabei wurden – einem Vorschlag von Krumhansl (1990) folgend – die Tonhö-

henwerte des jeweils vorausgehenden und nachfolgenden Taktes mit je der halben Gewichtung miteinbezogen. Als Ergebnis der Berechnungen resultieren für einen Ähnlichkeitsvergleich zwei Folgen von Tonalitätssymbolen, wie z.B. F d F d a B und F F d d F d d. Als Ähnlichkeitsmaß wird hier die Edit Distance (s.o.) für den Vergleich zweier Symbolstrings dieser Form genutzt. Das verwendete Ähnlichkeitsmaß für den Vergleich der impliziten harmonischen Tonalitäten ist:

harmcore harmonische Korrelation nach Krumhansl, Vergleich der Tonalitätsstrings über Edit Distance

G) Melodien als Sets von musikanalytisch relevanten Parametern (Selfridge-Field):

Im Gegensatz zu den bis hierhin erläuterten Ähnlichkeitsmaßen beschränkt sich das Modell von Eleanor Selfridge-Field nicht auf einen einzigen Parameter von Melodien, sondern vergibt für den Grad der Übereinstimmung auf mehreren Parametern zwischen zwei Melodien Punkte. Es ist als scoring-Modell eigentlich für den Einsatz bei der unscharfen Suche in Musikdatenbanken konzipiert und weist in seiner ursprünglichen Konzeption auch Module für den Vergleich von mehrstimmigen harmonischen Begleitungen auf (Selfridge-Field, 2002 und persönliche Kommunikation 2003). Es ist damit ein eher ‚musikalisch‘ orientiertes Modell im Gegensatz zu den meisten bisherigen Ähnlichkeitsmaßen, deren Anwendungsbereich auch die in Zeichenketten darstellbaren Informationen anderer Domänen sein können. Für den Einsatz in dieser Arbeit zur Bestimmung der Ähnlichkeit von einstimmigen Melodien wurde das Modell geringfügig modifiziert bzw. reduziert.

Der Algorithmus des Modells ist in fünf Stufen gegliedert:

1. Bestimme das Metrum und die häufigste Tondauer der beiden Melodien. Wenn das Metrum nicht übereinstimmt vergib 0 Punkte. Wenn das Metrum übereinstimmt, aber nicht die häufigste Tondauer, vergib 0,5 Punkte. Wenn das Metrum und die häufigste Tondauer übereinstimmt, vergib 1 Punkt.
2. Bestimme für beide Melodien die Tonalitäten jedes Taktes mit Hilfe des Krumhansl-Algorithmus (s.o.). Bestimme je die häufigste Tonart für beide Melodien (Modus der Tonarten). Dies ist die Haupttonart der Melodie. Bestimme die Anzahl der Übereinstimmungen zwischen den Tonalitäten einzelnen Takte beider Melodien in ganzer Länge. Für Übereinstimmungen von weniger als 80% vergib 0 Punkte, für Übereinstimmungen von 80% bis 89% vergib 0,33 Punkte, für Übereinstim-

mungen von 90% bis 99% vergib 0,67 Punkte, für Übereinstimmungen von 100% vergib 1 Punkt.

3. Bestimme für beide Melodien für jeden Ton die Tonleiterstufe (scale degree) relativ zur Haupttonart der jeweiligen Melodie.
4. Bestimme die Anzahl der Übereinstimmungen zwischen den Tonleiterstufen auf den starken Zählzeiten (1 und 3 im 4/4-Takt) beider Melodien in ganzer Länge. Für Übereinstimmungen von weniger als 80% vergib 0 Punkte, für Übereinstimmungen von 80% bis 89% vergib 0,67 Punkte, für Übereinstimmungen von 90% bis 99% vergib 1,33 Punkte, für Übereinstimmungen von 100% vergib 2 Punkte.
5. Bestimme die Anzahl der Übereinstimmungen zwischen den Tonleiterstufen auf den schwachen Zählzeiten (2 und 4 im 4/4-Takt) beider Melodien in ganzer Länge. Für Übereinstimmungen von weniger als 80% vergib 0 Punkte, für Übereinstimmungen von 80% bis 89% vergib 0,33 Punkte, für Übereinstimmungen von 90% bis 99% vergib 0,67 Punkte, für Übereinstimmungen von 100% vergib 1 Punkt.

Das Ergebnis wird durch die maximale Punktzahl 5 geteilt.

Das Ähnlichkeitsmaß nach dem Modell von Selfridge-Field, das in dieser Arbeit Anwendung findet, ist das folgende:

esfmodk Scoring-Maß nach Selfridge-Field, basierend auf der Tonalitätsermittlung nach Krumhansl (s.o.).

4.2.2 Zusammenfassung Ähnlichkeitsmaße

Aus den 24 im vorangegangenen vorgestellten Ähnlichkeitsmaßen sollen in der zweiten Hälfte dieses Kapitels die Maße oder die Kombination von verschiedenen Maßen ermittelt werden, die die größte Übereinstimmung mit den Ähnlichkeitsurteilen von musikalischen Experten zeigen. Die 24 vorgestellten Maße unterscheiden sich – wie oben ausführlich besprochen – hinsichtlich der verschiedenen musikalischen Parameter (absolute Tonhöhen und Intervalle, melodische Kontur, Rhythmus, melodische Motive, Rhythmus, implizite harmonische Tonalität), hinsichtlich des Berechnungsverfahrens und der auf unterschiedliche Weise aufbereiteten Ausgangsdaten. Eine Zusammenfassung der in dieser Arbeit untersuchten Ähnlichkeitsmaße und deren im folgenden verwendeten Abkürzungen stellt folgende Tabelle dar:

diff	durchschnittliche Differenz der Intervallabstände zwischen beiden Melodien
rawpc	Korrelationskoeffizient nach Pearson-Bravais der Tonhöhenwerte
rawpcw	gewichteter Korrelationskoeffizient nach Pearson-Bravais der Tonhöhenwerte
rawpcst	auf den Wertebereich 0-1 transformiertes r durch Setzen der negativen Werte von r gleich 0
rawpcwst	auf den Wertebereich 0-1 transformiertes r durch Setzen der negativen Werte von r gleich 0, zusätzlich gewichtet
conspc	Korrelationskoeffizient nach Pearson der gewichteten Konturersatzreihe, konstruiert nach Steinbeck
conspcst	Korrelationskoeffizient nach Pearson der gewichteten Konturersatzreihe, konstruiert nach Steinbeck, auf den Wertebereich 0-1 transformiert
conpc	Korrelationskoeffizient nach Pearson der gewichteten Konturersatzreihe, konstruiert nach Müllensiefen und Frieler
conpcst	Korrelationskoeffizient nach Pearson der gewichteten Konturersatzreihe, konstruiert nach Müllensiefen und Frieler, auf den Wertebereich 0-1 transformiert
fourr	Korrelationskoeffizient nach Pearson der Amplituden der fouriertransformierten Tonhöhenränge
fourrst	Korrelationskoeffizient nach Pearson der Amplituden der fouriertransformierten Tonhöhenränge, auf den Wertebereich 0-1 transformiert
fourw	Korrelationskoeffizient nach Pearson der Amplituden der fouriertransformierten, gewichteten Tonhöhenränge
fourrwst	Korrelationskoeffizient nach Pearson der Amplituden der fouriertransformierten, gewichteten Tonhöhenränge, auf den Wertebereich 0-1 transformiert
fourri	Durchschnittliche Differenz der Amplituden der fouriertransformierten Tonhöhenränge
rawed	Edit Distance der (transponierten) Tonhöhenwerte
rawedw	Edit Distance der (transponierten) gewichteten Tonhöhenwerte
coned	Edit Distance der gewichteten Konturersatzreihen, konstruiert nach Müllensiefen und Frieler
consed	Edit Distance der gewichteten Konturersatzreihen, konstruiert nach Steinbeck
diffed	Edit Distance der Intervallfolgen
rhythfuzz	Edit Distance der zu fünf Klassen zusammengefassten Tondauern
rhythgaus	Kreuzkorrelation der ‚gaußifizierten‘ Onsetwerte
ngrsumco	Sum Common-Maß auf Basis der Intervalle
ngrukkon	Ukkonen-Maß auf Basis der Intervalle
ngrcoord	Coordinate matching (count distinct)-Maß auf Basis der Intervalle
ngrsumcr	Sum Common-Maß auf Basis der Intervallrichtungen (Parsons Code)
ngrukkor	Ukkonen-Maß auf Basis der Intervallrichtungen (Parsons Code)
ngrcoorr	Coordinate matching (count distinct)-Maß auf Basis der Intervallrichtungen

	(Parsons Code)
ngrsumcof	Sum Common-Maß auf Basis der Intervallklassen
ngrukkof	Ukkonen-Maß auf Basis der Intervallklassen
ngrcoorf	Coordinate matching (count distinct)-Maß auf Basis der Intervallklassen
ngrsumfr	Sum Common-Maß auf Basis der Dauernklassen
ngrukkfr	Ukkonen-Maß auf Basis der Dauernklassen
ngrcoofr	Coordinate matching (count distinct)-Maß auf Basis der Dauernklassen
harmcore	harmonische Korrelation nach Krumhansl, Vergleich der Tonalitätsstrings über Edit Distance
esfmodk	Scoring-Maß nach Selfridge-Field, basierend auf der Tonalitätsermittlung nach Krumhansl

Tabelle 4-8: Für den Vergleich mit den menschlichen Ähnlichkeitsurteilen verwendete algorithmische Ähnlichkeitsmaße.

Alle Maße wurden in einem einzigen *Softwaretool* implementiert, welches in seiner Kommandozeilenversion den Namen *Antiplagiator* und in seiner Windows-Version mit grafischem Interface die Bezeichnung *SIMILE* trägt. Die Programmierung des *Softwaretools* wurde von Klaus Frieler durchgeführt.

4.3 Erhebung subjektiver Ähnlichkeitsurteile

4.3.1 Das Hauptexperiment

Wie gezeigt kommt jedes der oben beschriebenen Ähnlichkeitsmodelle zu teilweise stark verschiedenen Ergebnissen bei der Berechnung der Ähnlichkeit zweier melodischer Gebilde. Dabei ist jedoch unklar, welches Modell für die Analyse und Modellierung von melodischen Gedächtnisprozessen, die das zentrale Thema im weiteren Verlauf dieser Arbeit ist, am besten geeignet ist. Der zu Anfang dieses Kapitels skizzierte Ansatz geht davon aus, dass dasjenige Modell, welches die größte Übereinstimmung mit Ähnlichkeitsurteilen menschlicher Experten zeigt, als optimales Modell gewählt werden sollte.

Zu diesem Zweck ist es nötig, Musikexperten und die verschiedenen Algorithmen die Versuchsmaterialien beurteilen zu lassen. Dazu wurden ein Haupt- und zwei Kontrollexperimente durchgeführt, die folgend detailliert beschrieben werden.

4.3.1.1 Hypothesen zum Zusammenhang der menschlichen Ähnlichkeitsurteile mit den experimentellen Variablen

Prinzipiell hat das Hauptexperiment einen explorativen Charakter. So ist sein vordringlichster Zweck das Erheben von menschlichen Ähnlichkeitsurteilen, die zur Beurteilung

der verschiedenen Ähnlichkeitsalgorithmen dienen sollen. Dennoch lassen sich für die im nachfolgend beschriebenen Design erhobenen Variablen auch inhaltliche Hypothesen formulieren, durch deren Bestätigung zusätzliche Erkenntnisse über die Prozesse gewonnen werden können, die der menschlichen Ähnlichkeitsbeurteilung von Melodien zugrunde liegen. Die Hypothesen sind als inhaltliche Alternativhypothesen formuliert, für die einzelnen statistischen Prüfungen wurden entsprechende Null-Hypothesen formuliert bzw. die entsprechenden Operationalisierungen für die jeweils adäquaten statistischen Tests vorgenommen:

Hypothese 1:

Die Ähnlichkeitsurteile der selektierten Musikexperten weisen einen starken Zusammenhang auf.

Hypothese 2:

Zwischen den Testkriterien Stabilität und Reliabilität des Ähnlichkeitsurteils und der musikalischen Expertise der Versuchspersonen besteht ein starker Zusammenhang.

Hypothese 3a:

Es besteht ein starker Zusammenhang zwischen der Anzahl der Abweichungen (Fehler) in zwei zu vergleichenden Melodien und der Ähnlichkeitsbeurteilung durch die selektierten Experten.

Hypothese 3b:

Es bestehen deutliche Unterschiede zwischen dem Einfluss, den die Abweichungen (Fehler) in den verschiedenen Dimensionen (Intervalle, Kontur, Rhythmus, ‚Modulation‘, Phrasenreihenfolge) auf das Ähnlichkeitsurteil der selektierten Experten haben.

Hypothese 3c:

Es besteht ein Zusammenhang zwischen der Position der Abweichungen (Fehler) im Gesamtverlauf der Melodien und dem Ähnlichkeitsurteil der selektierten Musikexperten.

Hypothese 4:

Es bestehen signifikante Unterschiede zwischen den Ähnlichkeitsurteilen für verschiedene Melodien unabhängig von den Abweichungen der einzelnen Melodievarianten. Diese unterschiedlichen Ähnlichkeitsurteile liegen in der unterschiedlichen Komplexität der Melodien begründet, die die Ähnlichkeitsurteile erschweren und damit beeinflussen.

4.3.1.2 Methode

Versuchspersonen:

Insgesamt nahmen 82 Studierende der Musikwissenschaft (Haupt- und Nebenfach) in zusammen vier Gruppentestsessions an dem Experiment teil. Unter den Teilnehmern wurden zehn CD-Einkaufsgutscheine im Wert von insgesamt 170,- Euro zur Motivationssteigerung verlost. Alle Teilnehmer besaßen eine musikalische Vorbildung (Instrumentalspiel für mindestens 2 Jahre). Die Qualität und der Grad dieser Vorbildung wurden über einen ausführlichen Fragebogen erhoben und wiesen zum Teil erhebliche Schwankungen auf.

Versuchsmaterialien:

Als Ausgangsmaterial wurden 14 Melodien überwiegend aus dem Bereich der populären Musik der letzten 35 Jahre sowie zwei Volksmelodien gewählt (s. Anhang). Alle Melodien waren Gesangsmelodien, hatten in etwa eine Länge von acht Takten, beinhalteten mehrere Phrasen mit mindestens zwei verschiedenen melodischen Motiven und zeichneten sich durch melodische Bewegung aus. Tonrepetitionen, wie in Strophen von Rock- und Popmusik üblich, wurden weitestgehend vermieden. Da das zu ermittelnde Ähnlichkeitsmodell zuerst für die Analyse von Gedächtnisfehlern eingesetzt werden soll, wurden zu jeder der 14 Melodien sechs Varianten konstruiert, insgesamt also 84 Varianten. Die Regeln für die Konstruktion der Varianten reflektieren typische Gedächtnisfehler, wie sie bei der menschlichen Wiedergabe von Melodien aus dem Gedächtnis auftreten. Als Heuristik für die Definition der Fehlertypen und deren Häufigkeiten dienten die Veröffentlichungen von Sloboda und Parker (1985), Oura und Hatano (1988), Zielinska und Miklaszewski (1992), McNab et al. (1996), Meek und Birmingham (2002) und Pauws (2002). Fünf Fehlertypen wurden mit jeweils drei Ausprägungsgraden und den ihren zugehörigen Auftrittswahrscheinlichkeiten definiert. Innerhalb eines Fehlertypus waren die jeweils drei Ausprägungsgrade gleich wahrscheinlich:

- Rhythmusfehler mit insgesamt $p=0,6$; $p(0 \text{ Rhythmusfehler}) = 0,4$; $p(3) = 0,2$; $p(6) = 0,2$; $p(9) = 0,2$;
- Tonhöhenfehler ohne Konturänderung (Stufen: 3, 6, 9 Fehler), insgesamt $p=0,4$;
- Tonhöhenfehler mit Konturänderung (Stufen: 3, 6, 9 Fehler), insgesamt $p=0,2$;
- Fehler bei der Phrasenreihenfolge (Stufen: 1, 2, 3 Fehler), insgesamt $p=0,2$;
- Modulationsfehler (d.h. ‚Verrutschen‘ in andere Tonart im Verlauf der Melodie), (Stufen: 1, 2, 3 Fehler), insgesamt $p=0,2$.

Für die Kombinationen aller möglicher Fehlerausprägungen wurden Wahrscheinlichkeiten errechnet und die Fehler so über die 84 Varianten verteilt. Die Tonhöhenfehlertypen b) und c) waren alternativ und konnten nicht in Kombination auftreten. Die Spannbreite der zufälligen Fehlerverteilung reichte von 0 bis 16 Fehlern.

Die Versuchstimuli wurden mit einem MIDI-Editor bearbeitet und anschließend im Grand Piano-Sound in das normale PCM-Audioformat verwandelt. Die Dauer der Melodien bzw. Varianten bewegte sich zwischen 12 und 20 Sekunden.

Als Beispiel für eine Referenzmelodie (Referenzmelodie D) und eine mit Fehlern versetzte Variante sind hier die Refrainmelodie des Dancetitels „Wonderland“ (Interpreten: Passion Fruit) und deren Variante D1 in der Notenschrift dargestellt:



Abb. 4-4: Originalmelodie D, Refrain des Titels „Wonderland“ (Interpreten: Passion Fruit).



Abb. 4-5: Konstruierte Variante D1 zur Originalmelodie D, Refrain des Titels “Wonderland”.

Die Variante D1 enthält im Vergleich mit der Referenzmelodie drei Abweichungen in der Rhythmusdimension (Tonwiederholungen und –auslassungen werden als Rhythmusfehler gezählt) und 9 Abweichungen bei der Tonhöhenkontur, die hauptsächlich in den Takten 7 und 8 zu finden sind.

Versuchsprozedur:

Die Versuchspersonen wurden zu Beginn jeder Sitzung über den Ablauf des Versuchsdurchgangs instruiert. Die Melodien und ihre zugehörigen Varianten wurden in Blöcken über eine geeignete Audio-Anlage präsentiert. Jede Melodie wurde zunächst in ihrer Originalgestalt präsentiert, um die Versuchspersonen mit der Melodie bekannt zu machen. Denn in einem Pretest hatten sich Lerneffekte im Verlauf eines Melodieblockes gezeigt, die durch das Bekanntmachen mit der Originalmelodie gemindert werden sollten. Nach einer 4-sekündigen Pause folgten in jedem Melodieblock sechs Paare von Originalmelodie und einer Variante. Die Paare waren jeweils durch 4-sekündige Pausen getrennt. Die Reihenfolge der Varianten war zufällig. Auf jeden Melodieblock folgte eine Pause von 20 Sekunden, in denen die Versuchspersonen auf einem Fragebogen anzugeben hatten, ob ihnen die Originalmelodie bekannt war und gegebenenfalls den Titel des Liedes. Je nach Session folgten drei bis fünf Melodieblöcke aufeinander. Die Dauer der Testsessions variierte entsprechend zwischen ca. 15 und 20 Minuten (ohne Versuchsinstruktionen).

Die Versuchspersonen beurteilten die Ähnlichkeit jedes Paares von Originalmelodie und Variante auf einer siebenstufigen Skala. Sie waren dabei angehalten sich vorzustellen, dass die Variante eine meist fehlerhafte Wiedergabe der Originalmelodie durch einen Musik-

schüler sei, die sie mit einer Punktzahl von 1 (geringe Ähnlichkeit) bis 7 (maximale Ähnlichkeit, Identität) auf einem Fragebogen bewerten sollten. Durch Verwendung dieser Hilfsvorstellung waren die Versuchspersonen ohne Schwierigkeiten in der Lage, die Ähnlichkeitsbeurteilungen vorzunehmen.

Nach der eigentlichen Testsession hatten die Versuchspersonen einen Fragebogen zu ihren gegenwärtigen und vergangenen musikalischen Aktivitäten auszufüllen. Der Fragebogen orientierte sich an dem von Mainz und Salthouse (1998) verwendeten (s. Anhang).

Design:

Da Ähnlichkeitsurteile von Musikexperten erhoben werden sollten, musste jede Versuchsperson zwei Voraussetzungen erfüllen, damit ihre Daten in die statistische Auswertung eingehen konnten: Die Beurteilung der Ähnlichkeit eines Melodiepaares sollte über eine gewisse Zeit in etwa konstant sein, und die Versuchspersonen sollten in der Lage sein, die Identität von Originalmelodie und Variante zu entdecken. Nur die Versuchspersonen, die diese Bedingungen der Urteils Konstanz und der Zuverlässigkeit in Hinsicht auf identische Melodien erfüllten, wurden aus der Gesamtheit der 82 Teilnehmer als ‚Musikexperten‘ ausgewählt und ihre Daten entsprechend weiterverarbeitet.

Zur Operationalisierung dieser beiden Kriterien wurde ein Test-Retest-Design verwendet, bei dem dieselbe Gruppe von Versuchspersonen eine Woche später die selbe Testprozedur durchlief, wobei neben vorher unbekannten Melodien ein Melodieblock der Vorwoche ein weiteres Mal beurteilt werden musste. Zwei verschiedene Seminare des musikwissenschaftlichen Instituts der Universität Hamburg wurden gemäß dieses Designs getestet. Den Seminaren wurden zweimal drei bzw. zweimal fünf Melodieblöcke zur Bewertung präsentiert.

Um das Kriterium der Urteils Konstanz zu erfüllen, wurden die zweimal sechs Beurteilungen einer Versuchsperson desselben Melodieblocks aus zwei verschiedenen Wochen mit dem τ_b -Maß nach Kendall verglichen, welches die Wahrscheinlichkeit wiedergibt, dass beide Rangreihen auf dieselbe Population zurückgehen. Der geforderte Zusammenhang musste mindestens 0,5 betragen.

Für das Kriterium der Identitätserkennung mussten mindestens 85% der getesteten Paare, die eine identische Variante enthielten, mit den Ähnlichkeitswerten 6 oder 7 beurteilt werden. Dies entsprach bei der ersten Seminargruppe dem Erkennen von vier aus insgesamt

vier identischen Varianten und bei der zweiten Gruppe sieben aus acht identischen Varianten.

Rechnerische Behandlung der erhobenen Daten:

Zur Aufbereitung für die weitere Verarbeitung und zur Prüfung der beiden Kriterien des Musikexpertentums wurden die erhobenen Daten in folgenden fünf Schritten ausgewertet:

1. Es wurden alle Vpn. ausgeschlossen, die nicht an Test- und Retestsession teilgenommen hatten. Es bleiben 22 von 57 bzw. 15 von 25 Vpn. in der Auswertung.
2. Es wurden alle Vpn. ausgeschlossen, die nicht 4 von 4 (bzw. 7 von 8) identischen Varianten mit mindestens dem Wert 6 beurteilt hatten. Es blieben 15 bzw. 14 Vpn. in der Auswertung.
3. Es wurden alle Vpn. ausgeschlossen, deren Beurteilungen der Melodie-Variante-paare von Melodie A (bzw. Melodie G) von Woche 1 zu Woche 2 einen geringeren Zusammenhang als 0,5 auf dem τ_b -Maß nach Kendall aufwiesen. Es blieben 9 bzw. 14 Vpn. in der Auswertung, also 23 insgesamt.
4. In einem Pretest, 14 Tage vor der eigentlichen Testserie mit neun Teilnehmern, zeigten die Vpn. die Tendenz, jeweils bei dem ersten Vergleichspaar einer Melodie im mittleren Bereich der Skala anzukreuzen, unabhängig davon, wie viele Fehler die Variante des ersten Vergleichspaares enthielt. Befragungen im Anschluß an den Pretest ergaben, dass die Vpn. zu Anfang jedes Blockes erst mindestens ein Vergleichspaar benötigten, um sich die Originalmelodie einzuprägen und so keine sicheren Urteile über die Ähnlichkeit von Originalmelodie und erster Variante abgeben konnten, was sie in der Mehrzahl veranlasste, die Mitte der Bewertungsskala zu wählen. Um diesen Effekt entgegenzuwirken, wurde im Haupttest vor den eigentlichen Vergleichsparen die Originalmelodie einmal alleine zum Einprägen in das Gedächtnis präsentiert. Zusätzlich wurden die Daten auf zwei verschiedenen Wegen auf diese Lerneffekte geprüft:

A) Es sollten alle frühen Durchgänge ausgeschlossen werden, deren Mittelwerte sich nicht signifikant unterscheiden. D.h. es wurde geprüft, ob die jeweils ersten Vergleichspare signifikant verschiedene Beurteilungsmittelwerte besaßen. Gegebenenfalls sollten auch die

zweiten oder dritten Vergleichspaare auf Mittelwertunterschiede geprüft werden. Die Tests auf Innersubjekteffekte innerhalb der Prozedur *Allgemeines Lineares Modell mit Messwiederholung* von SPSS ergaben mit $p < 0,000$ hoch signifikante Mittelwertsunterschiede bereits für die ersten Vergleichspaare. Im Pretest hatte sich kein signifikanter Unterschied zwischen den Mittelwerten der ersten Vergleichspaare ergeben ($p = 0,13$).

B) Es sollten die Streuungen der Bewertungen in den jeweils ersten Hälften der Melodieblöcke (Vergleichspaare 1-3) nicht signifikant geringer sein als die Streuungen in den zweiten Hälften (Vergleichspaare 4-6). Dazu wurden die Varianzen der ersten Hälften mittels des Levene-Tests mit denen der zweiten Hälften verglichen. Bei 10 der insgesamt 16 Melodieblöcke unterschieden sich die Varianzen der beiden Hälften signifikant: Sechsmal war die Varianz der ersten Hälfte geringer, viermal größer. Die Wahrscheinlichkeit für diese Verteilung unter der Voraussetzung, dass beide Ereignisse (Varianz in 1. Hälfte größer, Varianz in 2. Hälfte größer) gleich wahrscheinlich sind, lässt sich mit der bekannten Formel für binomiale Stochastiken errechnen:

$$P(k) = \binom{n}{k} p^k q^{n-k}$$

Dabei steht n für die Anzahl der Ereignisse (hier 10 Melodieblöcke), k für die Anzahl der Melodieblöcke, bei denen die Varianz der 1. Hälfte größer ist und $p = q$ für die angenommene Wahrscheinlichkeit von 0,5.

Die Wahrscheinlichkeit für die empirisch gefundene Verteilung von 6:4 ist demnach 0,205 unter der Voraussetzung, dass beide Ereignisse gleich wahrscheinlich sind. Dieser Wert stellt also keine Irrtumswahrscheinlichkeit dar, nach der die Hypothese, dass die Varianz in beiden Hälften der Melodieblöcke gleich verteilt wäre, zu verwerfen wäre.

Reihenfolge- oder Lerneffekte können also in den Beurteilungen der selektierten Musikexperten ausgeschlossen werden.

4.3.1.3 Ergebnisse

4.3.1.3.1 Übereinstimmung der menschlichen Ähnlichkeitsurteile und musikalische Expertise der selektierten Versuchspersonen

Sinnvollerweise lässt sich das erhobene Datenmaterial nur dann verwenden, wenn zwischen den Beurteilungen der Versuchspersonen, d.h. der selektierten Musikexperten, ein starker Zusammenhang besteht. Nur wenn die Experten untereinander sehr ähnlich urteilen, lässt sich ein menschliches Ähnlichkeitsurteil als Modell annehmen und sinnvoll mit

den Ähnlichkeitsalgorithmen vergleichen. Ansonsten wären die menschlichen Ähnlichkeitsurteile hochgradig subjektiv und divergent und als Instrument für die Analyse von Gedächtnisfehlern nicht brauchbar.

Für die Prüfung des korrelativen Zusammenhangs zwischen einzelnen Variablen stellen die Prozeduren *Faktorenanalyse* und *Reliabilitätsanalyse* in SPSS mehrere Maße zur Verfügung:

- Das Maß für den globalen Zusammenhang in der Korrelationsmatrix der Urteile aller selektierten Experten untereinander nach *Kaiser-Meyer-Olkin* ergab für die beiden Seminargruppen die Werte 0,890 („recht gut“, nach Kaiser, vgl. Brosius (1998, S. 647)) bzw. 0,94 („fabelhaft“) für die Seminargruppe 2.
- Unter den Werten für die Zusammenhänge der einzelnen Variablen (hier Vpn.) nach dem sogenannten *Measure of Sampling Adequacy (MSA)* fand sich kein kleinerer Wert als 0,802 („recht gut“) bzw. keiner kleiner als 0,889 („recht gut“) in der zweiten Seminargruppe. 4 von 9 bzw. 13 von 14 Versuchspersonen urteilten mit einem Zusammenhangswert von $> 0,9$ („prächtig“).
- Schließlich prüft der Bartlett-Test auf Sphärizität die Nullhypothese, dass die Korrelationskoeffizienten in der Grundgesamtheit den Wert 0 haben. Mit einer Irrtumswahrscheinlichkeit von $p < 0,000$ konnte diese Nullhypothese in beiden getesteten Gruppen zurückgewiesen werden.
- Cronbach's α ist ein Reliabilitätskoeffizient, der nicht nur den korrelativen Zusammenhang zwischen einigen der untersuchten Variablen (hier= Vpn.) überprüft, sondern insgesamt angibt, wie gut alle Variablen eine latente unidimensionale Größe messen. Als latente unidimensionale Größe lässt sich im vorliegenden Fall das ‚wahre‘ Ähnlichkeitsurteil eines Musikexperten annehmen, von dem die Urteile der einzelnen selektierten Vpn. z.T. abweichen. Mit 0,962 bzw. 0,978 ergaben sich sehr hohe α -Werte für die beiden Seminargruppen; viel höher als beispielsweise die bei Lamont und Dibben (2001, S. 253) genannten α -Werte (um 0,84), die dort schon als hoch bezeichnet wurden. Die selektierten Vpn. wichen im hier betrachteten Fall vom ‚wahren‘ Ähnlichkeitsurteil also nur sehr geringfügig ab.

Alle vier Maße weisen also einen starken Zusammenhang zwischen den Ähnlichkeitsurteilen der selektierten Musikexperten aus. Die erste inhaltliche Hypothese des Hauptexperimentes, dass ein starker Zusammenhang zwischen den Ähnlichkeitsurteilen der selektierten Musikexperten besteht, ist somit bestätigt.

Mit diesem Resultat lassen sich Beurteilungsdaten der Versuchspersonen als Ergebnis eines homogenen menschlichen Ähnlichkeitsmodells für die Bewertung von melodischen Gebilden interpretieren. Es zeigt sich damit, dass alle die Versuchspersonen, die beide Expertisekriterien (Urteilst Konstanz und Identitätserkennung) erfüllen konnten, also Urteile mit interner Konsistenz abgaben, auch untereinander äußerst ähnlich urteilten, somit auch eine externe Konsistenz aufwiesen. Für die Beurteilung melodischer Ähnlichkeit scheint also – zumindest in diesem experimentellen Kontext - das Expertenurteil zu existieren. Der Vergleich und die Modellierung mit den algorithmischen Ähnlichkeitsmodellen erscheint also möglich und sinnvoll.

Der Terminus ‚Experten‘ scheint für die 23 durch die beiden Kriterien selektierten Versuchspersonen nicht nur aufgrund der Selektion angemessen. Denn die Daten zur vergangenen und gegenwärtigen musikalischen Aktivität dieser Versuchspersonen verweisen auf eine sehr intensive Beschäftigung mit Musik, wie die folgende Tabelle zeigt:

Deskriptive Statistik

	N	Mittelwert	Standardab weichung	Varianz
ALTER	23	23,22	3,813	14,542
SEX	23	1,57	,507	,257
GEHÖRBIL	23	,65	,487	,237
Gehör in Monaten	23	4,74	6,099	37,202
Gehör, akt. Phase in Std/W	23	1,543	2,0885	4,362
Gehör, Mom. in Std/W	23	,130	,3444	,119
Instrument in Jahren	23	12,52	5,534	30,625
Üben Mom. in Std/W	22	3,68	3,630	13,180
Musik Mom. in Std/W	23	9,48	10,067	101,352
Üben, akt. Phase in Std/W	22	12,00	7,819	61,143
Musik, akt. Phase in Std/W	23	21,35	13,412	179,874
Bez. Unt. in Monaten	23	71,91	65,266	4259,628
Bez. Auftr.	22	21,45	42,825	1833,974
Solo-Auftritte	21	14,81	32,876	1080,862
Begleitauftritte	21	7,71	10,321	106,514
Gruppenauftritte	23	29,96	46,491	2161,407
Blattspiel, Mom. in Std/W	23	,652	1,1816	1,396
Blattspiel, akt. Phase in Std./W	23	2,457	3,2680	10,680
Gültige Werte (Listenweise)	20			

Tabelle 4-9: Deskriptive Statistiken zum musikalischen Hintergrund der 23 ausgewählten Versuchspersonen (,Musikexperten') aus Experiment 1.

Die 23 Musikexperten spielen im Mittel seit 12,5 Jahren ein Instrument, musizierten in der musikalisch aktivsten Phase ihres Lebens 21 Stunden in der Woche, wobei 12 Stunden auf das reine Üben entfielen. Im Durchschnitt genossen sie 72 Monate bezahlten Instrumentalunterricht und absolvierten 30 Auftritte in einer Gruppe oder einem Ensemble. Im Schnitt wurden 21 der insgesamt bisher gespielten Auftritte (inkl. Solo- und Begleitauftritte) bezahlt. Typisch für Studierende der Musikwissenschaft ist die geringere musikalische Aktivität in der Gegenwart im Vergleich mit einer früheren – aktiveren – musikalischen Phase. Dies dürfte beispielsweise bei Studierenden an einer Musikhochschule grundsätzlich verschieden sein.

4.3.1.3.2 Zusammenhang zwischen musikalischer Expertise und der Erfüllung der Testkriterien

Wie die gerade dargestellten deskriptiven Statistiken nahe legen, scheinen große musikalische Aktivität in Vergangenheit und Gegenwart und das Erfüllen der beiden Testkriterien

zusammen aufzutreten. Es stellt sich damit die Frage, ob sich ein statistischer Zusammenhang zwischen musikalischer Expertise und der Erfüllung der Testkriterien finden lässt.

Zu diesem Zweck war es nötig, die Daten aus den insgesamt 16 Fragen des Fragebogens zu den musikalischen Aktivitäten auf eine oder wenige Dimensionen der musikalischen Expertise zu verdichten.

Als Vorbild diente hier – wie für den Fragebogen zur musikalischen Expertise – das Vorgehen von Meinz und Salthouse (1998), die aus den abgefragten Variablen durch Einsatz einer Faktorenanalyse eine Dimension der musikalischen Expertise ermitteln konnten und die Faktorwerte jeder Versuchsperson auf dieser Dimension zu weiteren statistischen Berechnungen verwendeten. Der Fragebogen erhob aus diesem Grund ausschließlich Anzahlen und Häufigkeiten bezüglich verschiedener Aspekte der musikalischen Aktivität. Die daraus generierten Variablen hatten so jeweils ein echtes Intervallskalenniveau, welches für die Berechnung der Korrelationen in einer Faktorenanalyse Voraussetzung ist.

Die Berechnung der Faktorenanalyse geschah in folgenden Schritten:

- Die Daten aller Vpn aus beiden Experimentalgruppen wurden in einer Datei zusammengeführt. Es resultierten 82 Fälle.
- Es wurden alle Vpn. ausgeschlossen, die bei mehr als vier der 16 Fragen zu ihrer musikalischen Aktivität keine Angaben gemacht hatten. Es blieben 78 Vpn. in der Auswertung.
- Die übrigen fehlenden Werte wurden durch die Mittelwerte der einzelnen Variablen ersetzt.
- Es wurden mehrere Faktorenanalysen mit der Gesamtzahl aller 16 Variablen und mit verschiedenen Auswahlen von Variablen gerechnet. Für die Auswahlen wurden einzelne oder mehrere Variablen ausgeschlossen, bei denen viele Vpn. den Wert 0 angegeben hatten (z.B. *Auftritte als Begleiter*, *Blattspiel*) und die deshalb als nicht besonders unterscheidungskräftig gelten konnten. Bei keiner dieser Faktorenanalysen ergab sich ein erster Faktor, der mehr als 31% der Gesamtvarianz aufklärte. Zudem zeigten die Maße für den korrelativen Zusammenhang der Variablen, das *Kaiser-Meyer-Olkin-Maß* (KMO) und das *Measure of Sampling Adequacy* (MSA), relativ schlechte Werte, die die Verwendung und Interpretation der faktorenanalytisch gewonnenen Daten für nicht ratsam erscheinen ließ.

- Da die größte musikalische Aktivität bei Studierenden der Musikwissenschaft in der Regel in der Vergangenheit (vor dem Studium) liegt, schien es inhaltlich geboten, Variablen auszuwählen, die vor allem die Beschäftigung mit Musik in einer zurückliegenden Phase größter musikalischen Aktivität abbilden, bzw. die musikalische Aktivität über das ganze bisherige Leben kumulieren. Folgende 8 Variablen wurden aufgrund dieser inhaltlichen Überlegung ausgewählt:
 - Instrumentalspiel (in Jahren),
 - Übestunden in aktivster Phase (in Stunden / Woche),
 - musikalische Aktivität in aktivster Phase (in Stunden / Woche),
 - bezahlter Unterricht (in Monaten),
 - bisherige Gehörbildung (in Monaten),
 - Gehörbildung in aktivster Phase (in Stunden / Woche),
 - Anzahl der gegen Bezahlung absolvierten Auftritte, Anzahl der Auftritte in einer Gruppe oder Ensemble.

Mit diesen 8 Variablen ließ sich eine Faktorenanalyse rechnen, bei der Maße für den korrelativen Zusammenhang zwar keine guten, aber insgesamt geeignete Werte zeigten: Nullhypothese in Bartlett-Test auf Sphärizität mit $p < 0,000$ zurückgewiesen, KMO: 0,574, MSA: Werte für die einzelnen Variablen zwischen 0,453 und 0,704.

Lediglich die ersten drei extrahierten Faktoren besaßen einen Eigenwert > 1 . Eine Beschränkung auf diese drei Faktoren schien auch nach Inspektion des Screeplots sinnvoll:

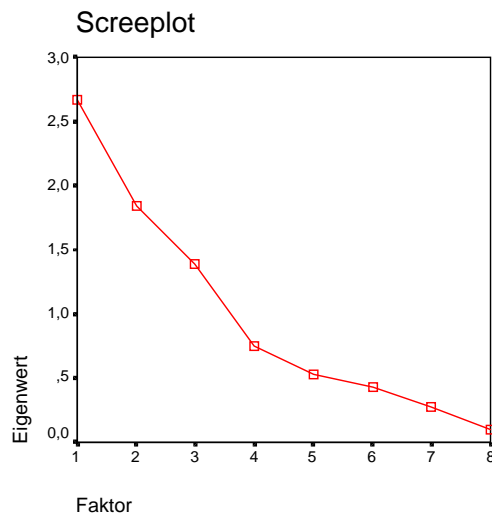


Abb. 4-6: *Screeplot* der Faktorenanalyse mit 8 ausgewählten Variablen zur musikalischen Aktivität der Vpn. Auch die Kommunalitäten der 8 verwendeten Variablen sprechen für eine gute Faktorlösung, da keine der Variablen nach der Faktorextraktion eine Kommunalität von weniger als 0,55 aufweist.

Kommunalitäten		
	Anfänglich	Extraktion
Instrument in Jahren	1,000	,746
Üben, akt. Phase in Std/W	1,000	,725
Musik, akt. Phase in Std/W	1,000	,757
Bez. Unt. in Monaten	1,000	,629
Bez. Auftr.	1,000	,936
Gehör in Monaten	1,000	,551
Gehör, akt. Phase in Std/W	1,000	,639
Gruppenauftritte	1,000	,927

Extraktionsmethode: Hauptkomponentenanalyse.

Tabelle 4-10: Kommunalitäten der acht an der FA beteiligten Variablen.

Der erste Faktor dieser Lösung konnte 33% der Gesamtvarianz aufklären. Die ersten drei Faktoren zusammen schon knapp 74%, wie aus folgender Tabelle ersichtlich ist.

Erklärte Gesamtvarianz

Komponente	Anfängliche Eigenwerte		
	Gesamt	% der Varianz	Kumulierte %
1	2,669	33,361	33,361
2	1,845	23,065	56,426
3	1,394	17,430	73,856
4	,754	9,429	83,286
5	,531	6,640	89,926
6	,430	5,371	95,297
7	,281	3,509	98,806
8	,096	1,194	100,000

Extraktionsmethode: Hauptkomponentenanalyse.

Tabelle 4-11: Erklärte Gesamtvarianz der Faktormodelle unterschiedlicher Dimensionalität.

Der erste der drei Faktoren lässt sich als *Musikalische Aktivität als Folge (formaler) musikalischer Ausbildung* interpretieren (s. Tabelle unten). Es laden hier das Instrumentalspiel und die Gehörbildung besonders hoch. Faktor 2 wäre dagegen die *Musikalische Praxis*, da hier deutlich die bezahlten und die Gruppenauftritte im Vordergrund stehen und bezahlter Unterricht sowie Gehörbildung negativ auf diesem Faktor laden. Der dritte Faktor steht für *Formale musikalische Ausbildung ohne besondere musikalische Aktivität*. Instrumentalspiel und bezahlter Unterricht laden hier hoch, das Üben und die musikalischen Aktivitäten laden dagegen stark negativ.

Komponentenmatrix^a

	Komponente		
	1	2	3
Instrument in Jahren	,680	-,161	,508
Üben, akt. Phase in Std/W	,563	-,037	-,638
Musik, akt. Phase in Std/W	,612	,095	-,612
Bez. Unt. in Monaten	,384	-,487	,494
Bez. Auftr.	,511	,810	,137
Gehör in Monaten	,587	-,453	-,024
Gehör, akt. Phase in Std/W	,660	-,449	-,045
Gruppenauftritte	,572	,713	,300

Extraktionsmethode: Hauptkomponentenanalyse.

a. 3 Komponenten extrahiert

Tabelle 4-12: Faktorladungen der acht Variablen auf den Faktoren der dreidimensionalen Lösung.

Zu jedem Fall (jeder Vpn.) wurden die Faktorenwerte der drei Faktoren als Variablen gespeichert. Diese Faktorenwerte, die Intervallskalenniveau aufweisen, wurden für die folgenden Analysen einzeln oder in Kombination als unabhängige Variablen verwendet.

Zusammenhang zwischen musikalischer Expertise und Erkennen identischer Varianten:

Um einen möglichen Zusammenhang zwischen der musikalischen Expertise in Form der Faktorwerte der drei extrahierten Faktoren und dem Erfüllen der Testkriterien aufzudecken, wurden eine Serie von Korrelationsrechnungen und inferenzstatistischen Tests vorgenommen.

Zunächst wurde für beide Teilnehmergruppen eine intervallskalierte Variable gebildet, die die Anzahl der erkannten identischen Varianten zählte und auf den Wertebereich zwischen 0 und 1 abbildete (Variable *identerk*). Diese Variable wurde zuerst mit den drei extrahierten Faktoren jeweils einzeln korreliert. Es ergaben sich Werte für den Korrelationskoeffizienten nach Pearson zwischen $-0,144$ (Faktor3) und $0,183$ (Faktor1). Keine der Korrelationen war statistisch signifikant.

Wie sich aus den Faktorwerten der Versuchspersonen auf den drei extrahierten Faktoren ersehen lässt, beschränkt sich die musikalische Expertise bei den meisten Versuchspersonen nicht auf einen Faktor, sondern die meisten Versuchspersonen zeigen hohe positive Werte auf mindestens zwei Faktoren: z.B. hohe Werte auf *Musikalische Aktivität als Folge (formaler) musikalischer Ausbildung* und auf *Musikalische Praxis*, was inhaltlich sehr gut nachzuvollziehen ist. Aus diesem Grund wurden die positiven Werte der einzelnen Faktoren addiert und die daraus resultierende Variable mit der Anzahl des richtigen Erkennens der identischen Variante (*identerk*) korreliert. Weder die aus den zwei Faktoren resultierenden Variablen ($F1+F2$, $F1+F3$, $F2+F3$) noch die aus allen drei Faktoren gebildete ($F1+F2+F3$), korrelierten hoch oder signifikant mit *identerk*; die Pearson-Koeffizienten bewegten sich zwischen $-0,08$ (FA2+3) und $0,1$ (FA1+2).

Es ließ sich also kein statistischer Zusammenhang zwischen der musikalischen Expertise der Versuchspersonen und ihren Leistungen beim Erkennen der identischen Varianten finden.

Zusammenhang zwischen musikalischer Expertise und der Urteilsstabilität:

Zunächst wurde mit einem t-Test geprüft, ob ein Zusammenhang zwischen musikalischer Expertise und der Stabilität der Ähnlichkeitsurteile zwischen Test- und Retestsession besteht. Dabei wurden die verschiedenen Faktorvariablen der musikalischen Expertise (s.o.) als abhängige und die Urteilsstabilität als abhängige Variablen eingesetzt, was inhaltlich

zwar dem vermuteten Prinzip von Ursache (Expertise) und Wirkung (stabiles Urteil) entgegengesetzt ist, aber angesichts des Messniveaus der Variablen (Intervallniveau bzw. dichotome Variablen) die einzige Möglichkeit für einen t-Test darstellt. Bestünde ein statistischer Zusammenhang zwischen beiden Variablen, so würde er sich gleichwohl durch diesen t-Test aufdecken lassen. Jedoch erbrachte kein einziger der insgesamt zehn durchgeführten t-Tests ein signifikantes Ergebnis (alle $p > 0,16$). Auch der zum t-Test analoge nicht-parametrische U-Test nach Mann und Whitney zeigte keinen einzigen signifikanten Unterschied an.

Einen hoch signifikanten Zusammenhang ergaben – wie zu erwarten war – sowohl t- als auch U-Test für das Erkennen der identischen Varianten (*identerk*) und das stabile Ähnlichkeitsurteil (t-test: $p = 0,0025$; U-Test: $p < 0,0015$).

Schließlich wurde noch in einem χ^2 -Test der Zusammenhang zwischen der Urteilsstabilität und einer dichotomen Variable der musikalischen Expertise geprüft. Für diese neue Variable wurden die positiven Faktorwerte der drei Faktoren addiert ($F1 + F2 + F3$) und die resultierenden Werte am Median der Variable dichotomisiert (zum Verfahren der Mediandichotomisierung s. Bortz, 1999, S. 550). Der χ^2 -Test für den Zusammenhang zwischen dieser Variable und der Urteilsstabilität ergab kein signifikantes Ergebnis ($p = 0,255$).

Musikalischer Expertise der selektierten Experten und der übrigen Versuchsteilnehmer:

Vergleicht man abschließend die Variablen der musikalischen Expertise bei den 23 durch die beiden Testkriterien selektierten Teilnehmern und den übrigen 59 Versuchspersonen bzw. den 14 Versuchspersonen, die an beiden Testsessions teilgenommen haben, so lassen sich ebenfalls keine signifikanten Unterschiede feststellen: Weder in den 10 t-Tests mit den Faktorenwerten der musikalischen Expertise als abhängigen Variablen sowie den Variablen, die sich durch Addition der Faktorenwerten ergeben, noch bei den 15 ‚Roh‘-Variablen des Fragebogens zu den einzelnen musikalischen Aktivitäten ergab sich auch nur ein signifikanter p-Wert; weder für den Unterschied zwischen den 23 Experten und den 59 übrigen Teilnehmern, noch für den Vergleich der 23 Experten mit den 14 ‚nicht-Experten‘, die ebenfalls an beiden Testsessions teilgenommen hatten. Die folgende Tabelle zu den musikalischen Aktivitäten der 59 nicht ausgewählten Teilnehmer enthält folglich auch sehr ähnliche Mittelwerte wie die oben dargestellte analoge Tabelle zu den selektierten Experten.

Deskriptive Statistik

	N	Mittelwert	Standardabweichung	Varianz
ALTER	54	24,07	2,854	8,145
SEX	57	1,63	,487	,237
GEHÖRBIL	57	,56	,501	,251
Gehör in Monaten	56	7,09	10,601	112,374
Gehör, akt. Phase in Std/W	57	1,368	1,8768	3,523
Gehör, Mom. in Std/W	57	,307	,8491	,721
Instrument in Jahren	57	12,35	5,759	33,160
Üben Mom. in Std/W	57	3,28	3,788	14,348
Musik Mom. in Std/W	57	7,46	7,091	50,288
Üben, akt. Phase in Std/W	55	13,24	9,879	97,591
Musik, akt. Phase in Std/W	56	19,30	13,831	191,306
Bez. Unt. in Monaten	53	70,30	65,963	4351,099
Bez. Auftr.	53	18,13	49,938	2493,809
Solo-Auftritte	54	7,81	15,373	236,342
Begleitauftritte	51	10,18	23,240	540,108
Gruppenauftritte	51	25,86	48,497	2352,001
Blattspiel, Mom. in Std/W	54	,852	2,0272	4,110
Blattspiel, akt. Phase in Std./W	53	3,236	3,5309	12,467
Gültige Werte (Listenweise)	44			

Tabelle 4-13: Deskriptive Statistiken zum musikalischer Hintergrund der 59 nicht ausgewählten Versuchspersonen aus Experiment 1.

Es ließ sich also in keinem einzigen Test ein Zusammenhang zwischen der musikalischen Expertise (gemessen durch die Faktorwerte der Faktorenanalyse) und den beiden Leistungskriterien hinsichtlich der Ähnlichkeitsurteile finden. Die inhaltliche Hypothese 2 ist also abzulehnen.

Da die musikalische Expertise sich jedoch in vielen vergleichbaren Testprozeduren als signifikant einflussreiche Größe erwiesen hat, liegt die Interpretation nahe, dass in der getesteten Stichprobe von Studierenden der Musikwissenschaft andere Größen, wie die momentane Aufmerksamkeit und Konzentrationsfähigkeit oder die Motivation, für die Leistung bei den Ähnlichkeitsbeurteilungen verantwortlich waren. Ab einem gewissen Niveau musikalischer Expertise, das für die getestete Stichprobe insgesamt gelten kann, scheint dieser Faktor seinen linearen Einfluss zumindest für die hier geforderte experimentelle Aufgabe zu verlieren.

4.3.1.3.3 Abhängigkeit der Ähnlichkeitsurteile von den verwendeten Fehlertypen

Von größtem Interesse für den späteren Vergleich der menschlichen Ähnlichkeitsurteile mit den verschiedenen algorithmischen Ähnlichkeitsmodellen ist der Einfluss, den die verschiedenen Fehlertypen und deren Ausprägung in den getesteten Melodievarianten auf die Einschätzung der Ähnlichkeit durch die 23 selektierten Experten ausüben. Auch in Hinsicht auf die in den beiden späteren Experimenten untersuchten Gedächtnisfehler ist die Bewertung der einzelnen Fehlerarten in diesem Experiment von Interesse. Fragen, die sich hier stellen, sind z.B. folgende: Werden kleine rhythmische Abweichungen bei der Rezeption wie bei der Reproduktion von Melodien großzügiger in Kauf genommen als Fehler bei melodischer Kontur und Intervallen? Wie wird das fälschliche Ausweichen in eine weit entfernte Tonart bewertet? Nimmt die Unähnlichkeit mit der Zahl der Fehler linear zu? Die folgende Analyse verfolgt dabei ähnliche Gedanken, wie die Untersuchung von McAdams und Matzkin (2001), die auch bei den Abweichungen zwischen zwei Melodien für die Modellierung von menschlichen Ähnlichkeitsurteilen ansetzt.

Eine bivariate Korrelation zwischen der Summe aller Fehler in den Melodievarianten und den Mittelwerten der Ähnlichkeitsbeurteilungen der Experten zeigte einen starken und signifikanten linearen Zusammenhang:

Korrelation nach Pearson-Bravais: $r = -0,727$; $p < 0,000$

Rangkorrelation nach Spearman: $\rho = -0,682$; $p < 0,000$

Mit der Anzahl der Fehler bzw. Abweichungen einer Variante von ihrem Original nimmt die Ähnlichkeit nach dem Urteil der Experten zwischen beiden ab bzw. die Unähnlichkeit zu. Die inhaltliche Hypothese 3a kann somit bestätigt werden.

Trotz dieser hohen und signifikanten Korrelationswerte scheint der Zusammenhang von Fehlerzahl und Ähnlichkeitsbeurteilung nicht perfekt linear zu sein, wie das folgende Diagramm verdeutlicht:

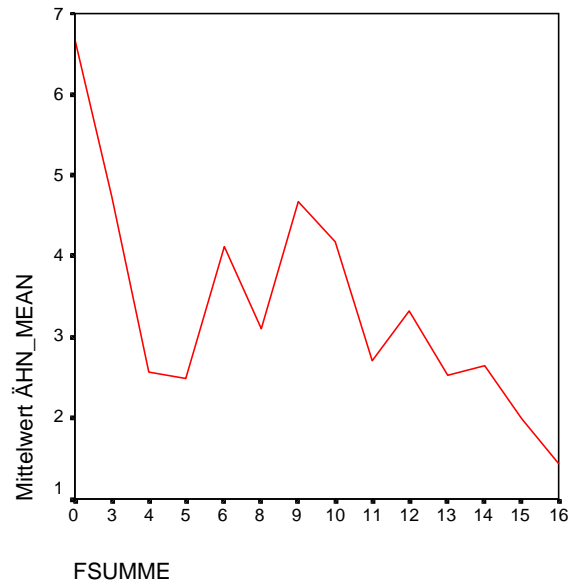


Abb. 4-7: Zusammenhang zwischen Fehlersumme und mittlerer Ähnlichkeit der 84 Melodievarianten

Varianten mit mittleren Fehlerzahlen (besonders 9 und 10) wurden offensichtlich ähnlicher beurteilt als Varianten, die sich nur durch 4 oder 5 Abweichungen von ihren Originalen unterschieden.

Gründe für diesen nur teilweise linearen Zusammenhang zwischen der Zahl der Abweichungen der Varianten und den Ähnlichkeitsbeurteilungen lassen sich in der unterschiedlichen Bedeutung, d.h. Gewichtung, der einzelnen Fehlervariablen, im Faktor Fehlerposition und in einem Störterm, der auf die urteilenden Versuchspersonen zurückgeht, vermuten. Diese einzelnen Faktoren sollen im folgenden untersucht werden.

Gewichte der Fehlervariablen:

Um die Gewichte der einzelnen Fehlervariablen für die Einschätzung der Ähnlichkeit durch die Experten zu ermitteln, wurde eine lineare Regression mit den fünf verwendeten Fehlerarten als Regressionsfaktoren und den Mittelwerten der Ähnlichkeitswerte der selektierten Experten als abhängige Variable gerechnet.

Wie aus den folgenden Tabellen zu erkennen ist, bewegt sich das Regressionsmodell insgesamt auf einem sehr guten Niveau: 79,9% der Streuung werden durch die fünf Prädiktorvariablen erklärt. Die Summe der erklärten quadrierten Abweichungen (166,9) übersteigt die Summe der residualen quadrierten Abweichungen (42) deutlich, weshalb das Modell in der ANOVA-Tabelle insgesamt einen hoch signifikanten Wert ($p < 0,001$) aus-

weist. Allerdings ist der Standardfehler des Schätzers mit 0,73379 noch recht hoch, sodass die beobachteten Werte durch die Regressionsgleichung noch deutlich überschätzt werden.

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,894 ^a	,799	,786	,73379

a. Einflußvariablen : (Konstante), FMODULAT, FREIHENF, FRHYTHM, FINTERVA, FKONTUR

ANOVA^b

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	166,609	5	33,322	61,886	,000 ^a
	Residuen	41,998	78	,538		
	Gesamt	208,607	83			

a. Einflußvariablen : (Konstante), FMODULAT, FREIHENF, FRHYTHM, FINTERVA, FKONTUR

b. Abhängige Variable: ÄHN_MEAN

Tabelle 4-14a und 4-14b: Werte zur Modellgüte der Regression von mittlerem Ähnlichkeitswert der Versuchspersonen auf Fehlerzahl in den verschiedenen Dimensionen.

Die β -Koeffizienten der fünf Prädiktorvariablen sind alle hochsignifikant und weisen das erwartete Vorzeichen auf (ein negativer Zusammenhang zwischen Fehlerzahl und Ähnlichkeit wurde erwartet), wie aus folgender Tabelle zu ersehen ist. Es scheint also keinen Grund zu geben, einen oder mehrere Faktoren aus der Regressionsanalyse auszuschließen.

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	6,126	,162		37,766	,000
	FRHYTHM	-,132	,023	-,294	-5,646	,000
	FINTERVA	-,283	,026	-,606	-11,049	,000
	FKONTUR	-,263	,031	-,466	-8,489	,000
	FREIHENF	-,821	,090	-,470	-9,167	,000
	FMODULAT	-,761	,095	-,414	-8,039	,000

a. Abhängige Variable: ÄHN_MEAN

Tabelle 4-15: Koeffizienten der Regression von mittlerem Ähnlichkeitswert der Versuchspersonen auf Fehlerzahl in den verschiedenen Dimensionen.

Die errechneten β -Gewichte lassen sich dahingehend interpretieren, dass rhythmische Fehler für die Einschätzung der Varianten als unähnlich eine geringere Bedeutung haben

als die übrigen Fehlertypen. Intervallfehler besitzen den größten Betrag unter den standardisierten β -Gewichten. Dies mag auch durch die größere Häufigkeit und den größeren *range* (0-9) der Intervallfehler bedingt sein. Denn betrachtet man die nicht-standardisierten Koeffizienten (1. Spalte der Tabelle) so zeigen dort Reihenfolge- und Modulationsfehler die größten Beträge, sodass diese als bedeutsamste Einflussgrößen für das Ähnlichkeitsurteil gelten können, wenn man von der absoluten Anzahl der Fehler ausgeht. Damit kann die inhaltliche Hypothese 3b bestätigt werden, nach der zumindest teilweise deutliche Unterschiede zwischen den Stärken der Einflüsse der fünf verschiedenen Fehlertypen bestehen.

Dieser starke Zusammenhang lässt sich auch grafisch im Liniendiagramm nachvollziehen:

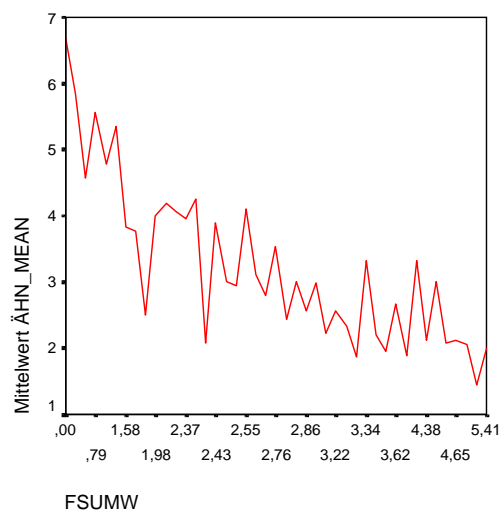


Abb. 4-8: Zusammenhang zwischen der an den Fehlerarten gewichteten Fehlersumme und der mittleren Ähnlichkeit der 84 Melodievarianten.

Die lineare Tendenz ist gut erkennbar, die Abweichungen der Ähnlichkeitswerte von der Regressionsgeraden, die sich durch die beobachteten Werte legen lässt, allerdings auch:

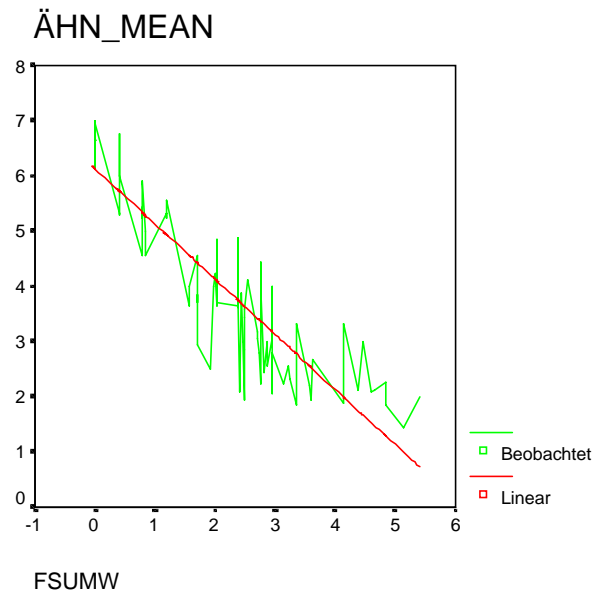


Abb. 4-9: Zusammenhang zwischen der an den Fehlerarten gewichteten Fehlersumme und der mittleren Ähnlichkeit der 84 Melodievarianten mit Gerade zum linearen Trend.

Dies lässt darauf schließen, dass noch ein oder mehrere Faktoren, die für das Ähnlichkeitsurteil der Experten mitbestimmend sind, in dem Regressionsmodell bisher nicht berücksichtigt wurden.

4.3.1.3.4 Fehlerposition:

Ein wahrscheinlicher Faktor ist die Position der Fehler im Verlauf der Variante. Denn wie beispielsweise Dewar et al. (1977) und Cuddy und Lyons (1981) beschrieben haben, kann das Entdecken von Abweichungen von der Position der Abweichung in einer melodischen Sequenz abhängen.

Zu vermuten ist, dass Fehler, die zu Beginn einer Variante zu hören sind, stärker das Ähnlichkeitsurteil beeinflussen als Fehler im weiteren Verlauf der Variante. Als Definition für den Beginn wurden zunächst die ersten vier Takte der Varianten definiert. Zur Überprüfung der Hypothese wurden insgesamt zehn neue Variablen gebildet, die die Anzahl der Fehler in der ersten Hälfte der Varianten (Takte 1-4) und der zweiten Hälfte (Takte 5-8 bzw. 9 oder 10 bei längeren Varianten) zählten.

Rechnet man eine lineare Regression mit diesen 10 Variablen, die die Fehlerzahl in den Takten 1-4 und 5-10 beschreiben, so erhält man ein Modell, dessen im Anhang dargestellte Maßzahlen eine leicht bessere Modellgüte anzeigen ($R=0,904$; korrigiertes $R^2=0,796$), als

das obige Modelle, welches nur die Fehlerzahlen ohne Berücksichtigung der Position als Prädiktorvariablen aufnahm.

Einzige Ausnahme bei der Operationalisierung waren die Reihenfolgefehler, die durch die Variable *freit14* ausschließlich in den Takten 1-4 gezählt wurden. Tatsächlich waren auch in vielen Varianten die Takte 5-10 von Reihenfolgefehlern betroffen. Da es die Natur der Reihenfolgefehler mit sich bringt, dass die Abweichungen einer Variante von ihrem Original bei einem Reihenfolgefehler immer mehrere Takte (i.d.R. im ersten und im zweiten Teil der Variante) betreffen, wurden die Reihenfolgefehler ausschließlich als Fehler der Takte 1-4 definiert. Denn tatsächlich existierte unter den Varianten keine einzige, die einen Reihenfolgefehler in der zweiten Hälfte aufwies, der nicht auch die erste Hälfte betroffen hätte.

Vergleicht man die β -Gewichte der einzelnen Fehlertypen für die erste und die zweite Hälfte der Varianten, so zeigen sich bei einigen Fehlertypen starke Unterschiede: Rhythmus- und Modulationsfehler tragen dann zu einer stärkeren Unähnlichkeitswahrnehmung bei, wenn sie am Anfang einer Melodievariante auftauchen. Konturfehler wurden in erster und zweiter Hälfte fast gleich bewertet, während die Intervallfehler deutlich stärker in der zweiten Hälfte die Ähnlichkeitsurteile beeinflussten. Das stärkste Gewichte besitzen die Reihenfolgefehler, die – wie oben beschrieben – nur in der ersten Hälfte definiert waren.

Durch die Konstruktion der Varianten bedingt weisen die Variablen eines Fehlertyps der ersten Hälfte mit den entsprechenden Variablen der zweiten Hälfte deutliche Korrelationen auf. Dieser Zusammenhang, der in der Literatur als Multikollinearität bezeichnet (Bortz, 1999, S. 438), kann die β -Gewichte der einzelnen Variablen verzerren und die obige Interpretation gefährden. Die in SPSS zur Verfügung stehenden Maße zur Prüfung auf Kollinearität im Regressionsmodell zeigten insgesamt jedoch keine kritischen Werte an (Toleranz aller Variablen $> 0,564$; Konditionsindex $< 4,78$).

Dieser enge Zusammenhang zwischen dem Ähnlichkeitsurteil der Experten und der an den Fehlertypen und Fehlerhälften gewichteten Fehlerzahlen lässt sich auch grafisch erkennen:

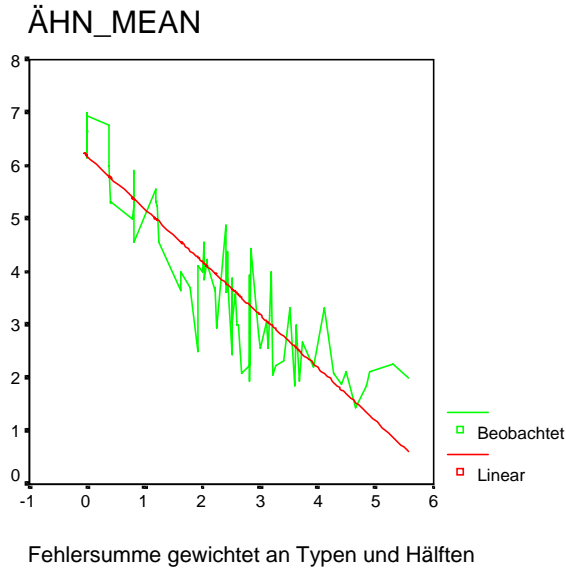


Abb. 4-10: Zusammenhang zwischen der an den Fehlerarten und an den Fehlerpositionen (Hälften) gewichteten Fehlersumme und der mittleren Ähnlichkeit der 84 Melodievarianten mit Gerade zum linearen Trend.

Die mittlere Fehlerposition:

Eine andere Beschreibung der Verteilung der Fehlerposition liefert die mittlere Fehlerposition. Die mittlere Fehlerposition kann man berechnen durch:

$$\bar{x}_f = \frac{\sum_{i=1}^n x_i f_i}{\sum f_i}$$

Dabei steht x_i mit $i \in \{1, \dots, n\}$ für die n Takte einer Melodie und f_i mit $f \in F = \{1, 2, 3, 6, 9\}$ für die Anzahl der Fehler in Takt i .

Für die fünf verschiedenen Fehlerarten (Rhythmus, Intervall, Kontur, Phrasenreihenfolge, Modulation) wurden jeweils getrennt die mittleren Fehlerpositionen für jede der 84 melodischen Varianten errechnet. Das Ergebnis sind die fünf neuen Variablen *fposrhy*, *fposint*, *fposkon*, *fposrei* und *fposmod*.

Mit diesen fünf neuen Variablen und den fünf Variablen, die Fehleranzahlen der verschiedenen Fehlertypen wiedergeben, wurde erneut ein lineares Regressionsmodell gerechnet, welches in den zugehörigen Tabellen im Anhang zusammengefasst ist.

Sowohl R (0,907) als auch das korrigierte R^2 (0,799) zeigen leicht bessere Werte als im vorangegangenen Regressionsmodell, welches nur die Information enthielt, in welcher Melodiehälfte sich wie viele Fehler befinden. Obwohl alle Positionsvariablen bis auf

fposrhy das erwartete Vorzeichen aufweisen, besitzen nur vier der fünf neuen Variablen signifikante β -Gewichte. Um Multikollinearität und Suppressioneffekte zu vermeiden, wurden die vier Variablen mit nicht signifikanten β -Gewichten aus der Regressionsanalyse entfernt. Die Ergebnisse dieser Analyse mit den fünf Variablen der Fehlerzahlen und der Variable zur Position der Intervallfehler (*fposint*) sind in den entsprechenden Tabellen im Anhang dargestellt.

Der multiple Korrelationskoeffizient R ist mit 0,901 etwas niedriger als in der Regression, die die Information der Fehlerhälften nutzte (0,904). Das korrigierte R^2 liegt mit 0,798 dagegen etwas über dem Wert der obigen Regression (0,796).

Die Grafik, die die Ähnlichkeitsmittelwerte und die an Position und Fehlertyp gewichteten Fehlerzahlen einander gegenüberstellt, verdeutlicht diesen engen Zusammenhang.

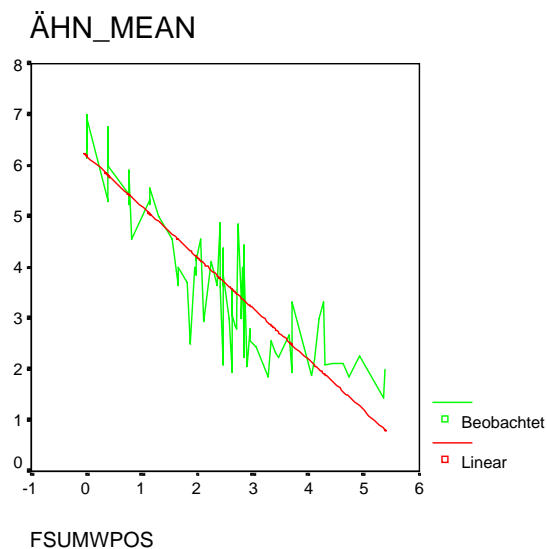


Abb. 4-11: Zusammenhang zwischen der an den Fehlerarten und an der mittleren Fehlerpositionen gewichteten Fehlersumme und der mittleren Ähnlichkeit der 84 Melodievarianten mit Gerade zum linearen Trend.

Zumindest die Position der Intervallfehler erweist sich damit als signifikante Einflussgröße für die Ähnlichkeitsurteile der Musikexperten. Je früher die Abweichungen in der Intervallstruktur im Verlauf einer Vergleichsmelodie auftauchen, desto geringer wird die Ähnlichkeit der beiden Melodien beurteilt. Die inhaltliche Hypothese 3c kann also für die Position der Intervallfehler bestätigt werden. Für die Position der Fehler der übrigen Typen jedoch nicht.

In der Interpretation scheint dies plausibel, wenn man auf der einen Seite die geringe Bedeutung der Rhythmusfehler und die große Bedeutung von Kontur-, Phrasenreihenfolge- und Modulationsfehlern bedenkt, wie sie durch die nicht-standardisierten Koeffizienten repräsentiert werden. Da Vertauschungen der Phrasenreihenfolge, Kontur- und Modulati-

onsfehler in der Regel sehr starke und gut erkennbare Abweichungen darstellen, können sie ein Ähnlichkeitsurteil auch dann genau so stark beeinflussen, wenn sie erst im späteren Verlauf einer Vergleichsmelodie auftauchen. Die Intervallfehler als ‚schwächere‘ Fehler haben einen umso stärkeren Einfluss auf das Ähnlichkeitsurteil der Experten, je näher sie dem vermutlich wahrnehmungsmäßig priorisierten Melodieanfang sind.

Die Fehlerdichte:

Einen Einfluß auf das Gewicht der mittleren Fehlerposition als Erklärungsfaktor mag die letzte in diesem Zusammenhang betrachtete Größe haben, die Fehlerdichte.

Die Überlegung, die hinter der Einbeziehung als Größe steht, ist die, dass die Fehlerposition vor allem dann für das Ähnlichkeitsurteil der Versuchspersonen von Bedeutung ist, wenn die Fehler sich über wenige Takte akkumulieren, d.h. quasi ‚zusammenklumpen‘. Nur in diesem Fall hat die Größe Fehlerposition einen Sinn, in dem sie angibt, wo die Fehler stattfinden.

Die Fehlerdichte wird nach der folgenden Formel berechnet:

$$d = \frac{\sum_{i=1}^n f_i}{T_{\max} - (T_0 + 1)}$$

Dabei steht f_i für die Anzahl der Fehler in Takt i , T_{\max} für den letzten Takt einer Variante, der einen Fehler enthält und T_0 für den ersten Takt einer Variante, der einen Fehler enthält. Die Anzahl der Fehler wird also durch die Spanne der Takte geteilt, auf die sich die Fehler einer Variante verteilen.

Mit Hilfe dieser Formel wurden drei neue Variablen (*fdichrhy*, *fdichint*, *fdichkon*) berechnet, die die Fehlerdichte für Rhythmus-, Intervall- und Konturfehler wiedergeben. Da die Modulations- und Reihenfolgefehler einer Variante nicht unabhängig voneinander verteilt wurden, ist die Berechnung der Dichte für diese Fehlerarten nicht sinnvoll.

Diese drei neuen Variablen wurden nun als Kehrwerte dazu verwendet, um die bestehenden Variablen der Fehlerposition zu gewichten und zwar nach dem Schema:

$$fposw = fpos \cdot \left(\frac{1}{fdich}\right)$$

Eine größere Fehlerdichte hat also eine niedrigere mittlere gewichtete Fehlerposition zur Folge, denn Fehler zu Beginn zeigten ja ein größeres Gewicht für die (Un-) Ähnlichkeitsbeurteilung der Versuchspersonen.

In der Implementierung der Formel lautete der Nenner ($fdich+0,0000001$), um bei Varianten mit 0 Fehlern einen nicht berechenbaren Term mit einer 0 im Nenner zu vermeiden. Diese Konstante verlor sich jedoch bei der Rundung der Variable $fposw$ auf zwei Dezimalstellen.

Anschließend wurde eine Regression mit den fünf Variablen der Fehlerzahlen pro Fehlertyp und den drei neuen gewichteten Fehlerpositionsvariablen gerechnet. Dabei ergab sich das in den zugehörigen Tabellen des Anhangs dargestellte Modell als optimales Modell, bei dem nur die an der Dichte gewichtete Fehlerposition als signifikanter Regressor unter den neu gebildeten Variablen in das Modell einging:

Da sowohl dieses Modell insgesamt eine bessere Modellgüte besitzt (korrigiertes R^2 jetzt: 0,808; vorher: 0,798) als auch der Signifikanzwert und der Standardfehler der gewichteten Fehlerposition deutlich geringer sind (jetzt: 0,002 bzw. 0,046; vorher: 0,021 bzw. 0,66) als beim vorigen Modell, bei dem die Dichte der Intervallfehler nicht einging, ist das neue Modell als bisher optimales zu akzeptieren.

Die beste Regressionsgleichung für die Schätzung der Ähnlichkeitsurteile der Musikexperten unter Einbeziehung der Größen Fehlerzahl, Fehlertyp, mittlere Fehlerposition und Fehlerdichte lautet also:

$$\hat{y} = 6,23 - 0,128*frhythm - 0,222*finterva - 0,278*fkontur - 0,867*freihenf - 0,744*fmodulat - 0,147*fposintw$$

Die Schätzungen nach dieser Gleichung korrelieren zu 0,907 signifikant ($p < 0,001$) mit den Ähnlichkeitsurteilen der Versuchspersonen. 82,2% der Varianz werden durch dieses lineare Modell erklärt. Wie gut der Zusammenhang zwischen der empirischen und der errechneten Variable durch einen linearen Trend dargestellt wird, lässt sich aus folgender Grafik erkennen.

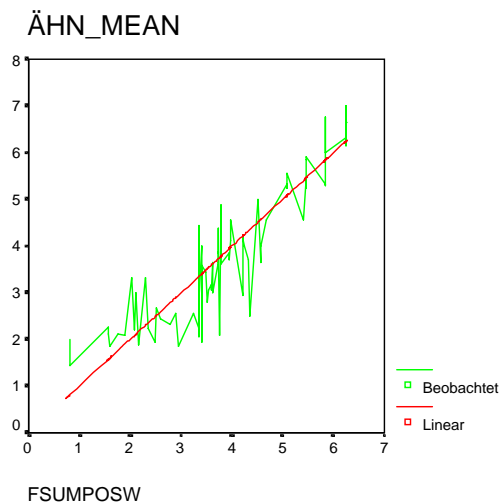


Abb. 4-12: Zusammenhang zwischen der an den Fehlerarten, an den Fehlerpositionen (genau) und der Fehlerdichte gewichteten Fehlersumme und der mittleren Ähnlichkeit der 84 Melodievarianten mit Gerade zum linearen Trend.

Noch etwas besser wird der Zusammenhang zwischen den beiden Variablen, wenn man nicht einen linearen, sondern einen quadratischen Zusammenhang unterstellt. In diesem Fall werden sogar 86% der Varianz aufgeklärt.

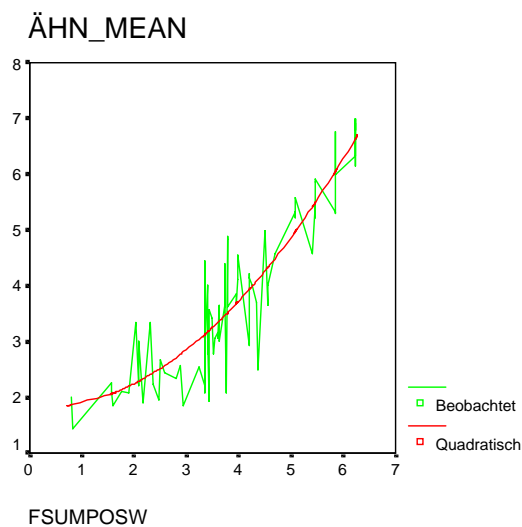


Abb. 4-13: Zusammenhang zwischen der an den Fehlerarten, an den Fehlerpositionen und der Fehlerdichte gewichteten Fehlersumme und der mittleren Ähnlichkeit der 84 Melodievarianten mit Kurve zum quadratischen Trend.

In jedem Fall lässt sich schlussfolgern, dass die getesteten Musikexperten ihr Ähnlichkeitsurteil von der Anzahl der Abweichungen – gewichtet nach Art der Abweichung und teilweise nach deren Position und Dichte im Melodieverlauf – abhängig machen. Es wird allerdings nicht unterstellt, dass dies als bewusster Prozeß bei den Musikexperten ablaufen muss.

4.3.1.3.6 Einfluß der unterschiedlichen Melodien

Die Fehler der Tonhöhe und des Rhythmus konnten in einer Studie von McAdams und Matzkin (2001, S. 71) nur einen Teil der Ähnlichkeitsbewertungen durch die Versuchspersonen erklären. Für den ungeklärten Teil der von ihnen erhobenen Werte vermuten die beiden Autoren eine Interaktion der Fehlertransformationen mit dem melodischen Material der Ausgangsmelodien. Das melodische Material an sich bzw. in Verbindung mit den speziellen Transformationen für das Experiment habe auch einen Einfluss auf die Ähnlichkeitsbewertungen.

Zur Prüfung des Einflusses der Struktur der Ausgangsmelodien in der vorliegenden Untersuchung wurde im ALM-Modul von SPSS eine Varianzanalyse (ANOVA) gerechnet, und zwar mit den mittleren Ähnlichkeitsbewertungen der Experten als abhängiger Variable und der Melodie (14 Stufen) als unabhängigem Faktor sowie der mehrfach gewichteten Fehlersumme (*fsumposw*: an Fehlertyp sowie Fehlerposition und Fehlerdichte der Intervalle) als intervallskaliertem Kovariable. Wie in der folgenden Tabelle ersichtlich, zeigte der Faktor *Melodie* einen signifikanten Haupteffekt jedoch keinen signifikanten Interaktionseffekt mit der Kovariable. Im ANOVA-Modell leistete die gewichtete Fehlersumme jedoch einen weit größeren Beitrag zum Modell (Quadratsumme des Effekts dieses Faktors: 142,5) als der Faktor Melodie (Quadratsumme 9,6), d.h. im Vergleich mit Fehlerzahl, -typ und Position und Dichte der Intervallfehler spielen die Eigenarten der unterschiedlichen Melodien eine sehr deutlich geringere Rolle für das Ähnlichkeitsurteil der Experten.

Tests der Zwischensubjekteffekte

Abhängige Variable: ÄHN_MEAN

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	188,466 ^a	27	6,980	19,408	,000
Konstanter Term	1,632	1	1,632	4,539	,038
FSUMPOSW	142,470	1	142,470	396,120	,000
MELODIE	9,555	13	,735	2,044	,033
MELODIE * FSUMPOSW	5,366	13	,413	1,148	,341
Fehler	20,141	56	,360		
Gesamt	1552,700	84			
Korrigierte Gesamtvariation	208,607	83			

a. R-Quadrat = ,903 (korrigiertes R-Quadrat = ,857)

Tabelle 4-16: Lineares Regressionsmodell zur Erklärung der mittleren Ähnlichkeitsurteile durch die Faktoren Melodie und Fehlersumme (an Fehlertypen, Fehlerposition und –dichte gewichtet, s. oben konstruiertes Modell).

Die deutlich geringere Bedeutung des Faktors Melodie als in der Studie von McAdams und Matzkin lässt sich wahrscheinlich durch das hier verwendete differenziertere Fehlersummenmodell erklären. Das „simplistic model“ (McAdams & Matzkin, 2001, S. 73) der beiden Autoren kennt nur zwei Transformations- bzw. Fehlertypen, nimmt keine Gewichtung der Fehlertypen vor, noch berücksichtigt es den Ort oder die Dichte des Fehlers. Durch die Berücksichtigung dieser zusätzlichen Komponenten können in der vorliegenden Studie die Ähnlichkeitsurteile der Musikexperten – bei McAdams waren die untersuchten Versuchspersonen Musiker und Nicht-Musiker – also weitgehend ohne den schwer beschreibbaren Einfluss des musikalischen Ausgangsmaterials erklärt werden.

Dennoch muss die inhaltliche Hypothese 4, die einen Einfluß der unterschiedlichen Melodien voraussagte, bestätigt werden.

4.3.2 Experiment 2 (Das Kontrollexperiment 1)

Zwei Gründe schienen Kontrollexperimente sinnvoll und nötig zu machen:

Zum einen sollten die oben beschriebenen Algorithmen zur Ähnlichkeitsbestimmung darauf hin geprüft werden, ob sie für Melodien, die in keinem offensichtlichen Zusammenhang stehen, eine entsprechend geringe Ähnlichkeit ausweisen, die nahe der maximalen Unähnlichkeit liegen sollte. Eine systematische Überschätzung der Ähnlichkeit sollte also abgeprüft werden. Dies gilt auch für die Musikexperten als Versuchspersonen: Werden Melodievarianten verschiedenen Ursprungs, d.h. von unterschiedlichen Ausgangsmelodien, deutlich unähnlicher beurteilt als Varianten, die derselben Ausgangsmelodie abstammen?

Zum anderen ist für die Bewertung der Ähnlichkeitsurteile der Musikexperten von Interesse, ob diese kontextabhängig sind, d.h. ob im Kontext von anderen Testmaterialien die Ähnlichkeitsurteile anders ausfallen. Denkbar ist hier eine lineare ‚Verschiebung‘ der Ähnlichkeitsurteile für Varianten der selben Melodien auf der siebenstufigen Skala hin zum Pol der maximalen Ähnlichkeit. Denn die Varianten anderer Melodien würden sich in Richtung des entgegengesetzten Pols der maximalen Unähnlichkeit gruppieren, sodass die siebenstufige Ähnlichkeitsskala anders ausgenutzt würde, die Ähnlichkeitswerte für die Varianten der selben Melodien jedoch nur monoton oder linear transformiert würden. Jedoch ist auch eine gänzlich andere Beurteilung der Varianten einer Melodie als in Experiment 1 denkbar, beispielsweise aufgrund der Transpositionen, die im Kontrollexperiment im Gegensatz zu Experiment 1 für alle Varianten vorgenommen wurden. Eine nicht-line-

are, vollkommen andere Bewertung der Ähnlichkeiten im Kontrollexperiment würde jedoch für die Modellierung der Ähnlichkeitsurteile der Experten gewisse Schwierigkeiten aufwerfen, da sie den aktuellen Kontext der gehörten bzw. verglichenen Melodien jeweils einbeziehen müsste, um reliabel zu sein. Alle beschriebenen algorithmischen Modelle liefern ja reliable, absolute Ähnlichkeitswerte, die von den zuvor verglichenen Melodien unabhängig sind.

4.3.2.1 Hypothesen

Aus diesen Überlegungen ergeben sich folgende drei Hypothesen, die als inhaltliche Alternativhypothesen formuliert sind (für die einzelnen statistischen Prüfungen wurden entsprechende Null-Hypothesen formuliert):

Hypothese 1:

Die Ähnlichkeitsbeurteilungen der Varianten, die nicht von der Referenzmelodie abstammen, sind signifikant geringer als die Beurteilungen von Varianten, die mit der Referenzmelodie verwandt sind.

Hypothese 2:

Die Ähnlichkeitsbewertungen der identischen Varianten aus dem Kontrolltest und dem Experiment 1 weichen im Mittel nicht signifikant voneinander ab.

Hypothese 3a:

Die Ähnlichkeitsbeurteilungen der identischen Varianten aus dem Kontrolltest und dem Experiment 1 weisen einen signifikanten Zusammenhang auf.

Hypothese 3b:

Die Ähnlichkeitsbeurteilungen der identischen Varianten aus dem Kontrolltest und dem Experiment 1 weisen in ihren Rangplätzen einen signifikanten Zusammenhang auf.

4.3.2.2 Methode

Versuchspersonen:

Teilnehmer des Experiments waren 16 Studierende der Musikwissenschaft des Hamburger Instituts, von denen elf in einer Gruppe und fünf jeweils in Einzelsitzungen die Versuchsprozedur durchliefen. Die Population, aus der die Versuchspersonen stammten, entspricht exakt der, aus denen sich die Teilnehmer des Experiments 1 rekrutierten.

Vier Teilnehmer hatten bereits 6 bzw. 9 Wochen zuvor an Experiment 1 teilgenommen. Durch diese Teilnahme waren sie zwar in der Lage, die Versuchsinstruktionen schneller zu verstehen, da sie fast identisch zu den Instruktionen des Experiments 1 waren. Aber durch den großen Zeitraum, der zwischen beiden Experimenten lag und durch die Transposition der Varianten im Kontrollexperiment kann es als sehr unwahrscheinlich gelten, dass diese Versuchspersonen einzelne Ähnlichkeitswerte aus dem Gedächtnis reproduzierten.

Versuchsmaterialien:

Für das Kontrollexperiment wurden dieselben Melodien und deren Varianten verwendet wie in Experiment 1. Es wurde jedoch jede Variante transponiert, um die absoluten Tonhöhen als ‚Gedächtniscue‘ auszuschließen. Die einzelnen Transpositionen wurden so vorgenommen, dass der Ambitus (höchster, tiefster Ton, Anfangstonhöhe) der Vergleichsvarianten mit dem Ambitus der Referenzmelodie weitmöglichst zur Deckung kam, die Transpositionen insgesamt aber keinem erkennbaren Muster in der Abfolge der Varianten nacheinander folgten.

Ausgewählt wurden zu jeder Referenzmelodie die sechs zugehörigen, ‚richtigen‘ Varianten (s. Experiment 1) sowie sechs weitere ‚falsche‘ Varianten, die von anderen Ursprungsmelodien abstammten und die gleichzeitig nach der Messung durch den Ähnlichkeitsalgorithmus des „Query-by-humming“-Verfahrens des Fraunhofer IDMT in Ilmenau eine bedeutsame Ähnlichkeit zu der jeweiligen Referenzmelodie zeigten, die aber nicht die originäre Ursprungsmelodie dieser Varianten war.

Die beiden gewählten Referenzmelodien waren die Melodie A (Demis Roussos: „Goodbye My Love, Goodbye“) und C (Wolfgang Petry: „Augen zu und durch“).

Prozedur:

Wie in Experiment 1 war es die Aufgabe der Versuchspersonen, die Ähnlichkeit von Melodievarianten mit einer gleichbleibenden Referenzmelodie auf einer siebenstufigen Ähn-

lichkeitsskala einzuschätzen. Wie in Experiment 1 hörten die Versuchspersonen zu Beginn eines jeden Durchlaufs die Referenzmelodie einmal ohne Vergleichsvariante, um Lerneffekte über den Versuchsdurchlauf hinweg auszuschließen.

Im Gegensatz zu Experiment 1 hatten die Versuchspersonen jedoch nicht sechs, sondern zwölf Vergleichspaare zu je einer Referenzmelodie zu beurteilen. In sechs der Vergleichspaare waren die transponierten Varianten der jeweiligen Referenzmelodie zu hören. In den übrigen sechs Vergleichspaaren bekamen die Versuchspersonen transponierte Varianten einer anderen Originalmelodie zum Vergleich angeboten (s. Abschnitt Versuchsmaterialien). Für diese Vergleichspaare wurde eine sehr geringe Ähnlichkeit erwartet, da hier Referenzmelodie und Variante auf den ersten Blick kaum oder gar keine Ähnlichkeiten zeigten.

Nach den *Ähnlichkeitsratings* hatten die Versuchspersonen denselben Fragebogen zu ihren musikalischen Aktivitäten auszufüllen, wie die Teilnehmer des Experimentes 1.

Design:

Die Versuchspersonen wurden in nur einer Session getestet. In dieser Session hatten sie zweimal zwölf Vergleichspaare zu bewerten und den abschließenden Fragebogen auszufüllen. Um zu ermitteln, welche Versuchspersonen ein stabiles Ähnlichkeitsurteil abgeben konnten, befanden sich unter den ersten zwölf Vergleichspaaren zwei identische Paare, die mit einer Abweichung nicht größer als 1 auf der siebenstufigen Skala beurteilt werden sollten. Zur Prüfung der Verlässlichkeit des Ähnlichkeitsurteils war in die Reihe der zweiten zwölf Vergleichspaare eine Variante eingefügt, die zwar transponiert, aber ansonsten identisch zur Referenzmelodie war. Diese identische Variante sollte mindestens mit dem Wert 6 auf der siebenstufigen Skala bewertet werden. Nur Versuchspersonen, die diese beiden Kriterien erfüllen, können als Musikexperten im Sinne der gestellten Aufgabe gelten, und nur die Ähnlichkeitsurteile dieser Experten sollten in die weitere Auswertung eingehen.

Rechnerische Behandlung der erhobenen Daten:

Zur Aufbereitung für die weitere Verarbeitung und zur Prüfung der beiden Kriterien des Musikexpertentums wurden die erhobenen Daten in folgenden Schritten ausgewertet:

- Es wurden alle Vpn. ausgeschlossen, die die identische Variante nicht mit mindestens dem Wert 6 beurteilt hatten. Es blieben alle 16 Vpn. in der Auswertung.
- Es wurden alle Vpn. ausgeschlossen, deren Beurteilungen der beiden identischen Melodie-Variantenpaare eine Differenz größer als 1 auswiesen. Es blieben 12 Musikexperten in der Auswertung.

Da sich in Experiment 1 keine auch nur annähernd signifikanten Lern- und Reihenfolgeeffekte gezeigt hatten, wurde auf die aufwendige Überprüfung dieser Effekte in den Daten des Kontrollexperiments verzichtet.

4.3.2.3 Ergebnisse

4.3.2.3.1 Übereinstimmung in den Urteilen der Musikexperten

Wie in Experiment 1 wurde der Zusammenhang zwischen den Urteilen der Musikexperten über die 24 getesteten Melodiepaare hinweg auf mehrfache Weise ermittelt:

- Das Maß für den globalen Zusammenhang in der Korrelationsmatrix nach *Kaiser-Meyer-Olkin* ergab für die Experten des Kontrollexperiments den Werte 0,811 („recht gut“, nach Kaiser, vgl. Brosius (1998, S. 647)).
- Unter den Werten für die Zusammenhänge der einzelnen Variablen (hier Vpn.) nach dem sogenannten *Measure of Sampling Adequacy (MSA)* fand sich kein kleinerer Wert als 0,696 („mäßig“). Die Urteile von 6 der 12 Vpn. ergaben einen Zusammenhangswert von $> 0,8$ („recht gut“).
- Der Bartlett-Test auf Sphärizität erbrachte mit einer Irrtumswahrscheinlichkeit von $p < 0,001$ keinen Anhaltspunkt dafür, dass keine signifikanten Korrelationen in der Grundgesamtheit vorlägen.
- Der Reliabilitätskoeffizient Cronbach's α erlangte einen hohen Wert von 0,9788, der darauf hinweist, dass die Urteile der einzelnen Versuchspersonen die unidimensionale Größe der wahren Ähnlichkeit insgesamt sehr gut messen.

Alle vier Maße weisen also einen deutlichen bis sehr starken Zusammenhang zwischen den Ähnlichkeitsurteilen der selektierten Musikexperten aus. Wie in Experiment 1 lassen sich Ähnlichkeitseinschätzungen der Versuchspersonen als Ergebnis eines homogenen menschlichen Ähnlichkeitsmodells für die Bewertung von melodischen Gebilden interpretieren. Auch die Ergebnisse dieses Kontrollexperiments weisen darauf hin, dass für die Beurteilung melodischer Ähnlichkeit das Expertenurteil existiert. Die hohe Übereinstim-

mung der Expertenurteile zeigte sich auch im Kontext sehr unterschiedlicher melodischer Varianten. Die intersubjektive Übereinstimmung scheint demnach nicht unter dem heterogenen Kontext der Testitems aus ‚richtigen‘ und ‚falschen‘ Varianten zu leiden, und eine Modellierung auch dieser Ähnlichkeitsurteile durch Algorithmen erscheint möglich. Die Ähnlichkeitswerte für weitere Berechnungen können demnach gemittelt werden.

Auch die zwölf ‚Experten‘ des Kontrollexperiments zeichnen sich durch einen hohen aktiven Umgang mit Musik aus. Dies lässt sich aus der folgenden Tabelle ablesen:

Deskriptive Statistik				
	N	Mittelwert	Standardabweichung	Varianz
ALTER	12	24,50	3,371	11,364
SEX	12	1,50	,522	,273
GEHÖRBIL	12	1,00	,000	,000
Gehör in Monaten	12	8,75	4,845	23,477
Gehör, akt. Phase in Std/W	12	2,500	1,7321	3,000
Gehör, Mom. in Std/W	12	,375	,7111	,506
Instrument in Jahren	12	14,58	3,450	11,902
Üben Mom. in Std/W	12	2,00	2,216	4,909
Musik Mom. in Std/W	12	3,75	3,621	13,114
Üben, akt. Phase in Std/W	10	13,60	8,356	69,822
Musik, akt. Phase in Std/W	11	16,27	9,023	81,418
Bez. Unt. in Monaten	11	122,18	51,772	2680,364
Bez. Auftr.	12	9,50	15,814	250,091
Solo-Auftritte	12	9,00	6,836	46,727
Begleitauftritte	12	19,92	23,975	574,811
Gruppenauftritte	12	31,58	25,069	628,447
Blattspiel, Mom. in Std/W	12	,833	1,5126	2,288
Blattspiel, akt. Phase in Std./W	12	4,500	3,2613	10,636
Gültige Werte (Listenweise)	8			

Tabelle 4-17: Deskriptive Statistiken zum musikalischen Hintergrund der zwölf selektierten Versuchspersonen („Experten“) aus Experiment 2.

Die selektierten Experten des Kontrollexperiments spielen durchschnittlich seit 14,6 Jahren ein Instrument, musizierten in der musikalisch aktivsten Phase ihres Lebens 16 Stunden in der Woche, wobei 13,6 Stunden auf das reine Üben entfielen. Im Mittel nahmen sie 122 Monate bezahlten Instrumentalunterricht und absolvierten 31,6 Auftritte in einer Gruppe oder einem Ensemble. Bezahlt wurden 9,5 der insgesamt bisher gespielten Auftritte (inkl. Solo- und Begleitauftritte). Diese Werte waren insgesamt sehr ähnlich zu denen aus Expe-

riment 1. Auch die typischerweise geringere musikalische Aktivität in der Gegenwart im Vergleich mit einer früheren aktiveren musikalischen Phase ist den Daten leicht zu entnehmen. Diese demographischen und aktivitätsbezogenen Daten verdeutlichen, dass die Teilnehmer des Kontrollexperiments und Versuchspersonen aus Experiment 1 aus derselben Population stammen und sich die Bezeichnung ‚Musikexperten verdient‘ haben.

Auf die aufwendige und detailreiche faktorenanalytische Verdichtung der Variablen der musikalischen Aktivität und der Zusammenhang mit dem Erfüllen der Testkriterien, wie sie in der Auswertung von Experiment 1 vorgenommen wurde, ist für das Kontrollexperiment verzichtet worden. Denn die Zahl der Versuchspersonen, die die beiden Kriterien nicht erfüllen konnten (0 Versuchspersonen, die die identische Variante nicht erkannten bzw. 4 Versuchspersonen, die nicht konstant urteilten), war für eine inferenzstatistische Auswertung zu gering.

4.3.2.3.2 Beurteilungsunterschiede zwischen ‚richtigen‘ Varianten und ‚falschen‘ Varianten anderer Melodien

Die Ähnlichkeitswerte von echten, ‚richtigen‘ Varianten der Referenzmelodie und ‚falschen‘ Varianten anderer Melodien im Vergleich mit der Referenzmelodie zeigten schon auf den ersten Blick starke Unterschiede: Innerhalb der beiden Testmelodien A und C wurde keine einzige ‚echte‘ Variante unähnlicher beurteilt als eine Variante anderer Melodien.

Zur Prüfung der Hypothese 1 wurde für jede Referenzmelodie getrennt jeweils ein t-Test für unabhängige Stichproben durchgeführt, bei dem die Unterschiede der Beurteilungsmittelwerte von Varianten mit der Referenzmelodie als Ursprung und solchen Varianten, die eine andere Melodie als Ursprung hatten, auf ihre statistische Signifikanz geprüft wurden. Die Mittelwerte der beiden Gruppen von 1,3 bei Melodie A bzw. 2,4 bei Melodie C (‚falsche‘ Varianten) und 3,9 bei Melodie A bzw. 5,1 bei Melodie C (‚richtige‘ Varianten) weisen deutliche Differenzen auf, die sich in den Teststatistiken jeweils auf dem 1-Pro-mille-Niveau ($p < 0,001$) als signifikant erwiesen.

Ein nicht-parametrischer U-Test nach Mann und Whitney, der auf den Rängen der beiden Experimentalgruppen beruht, erbrachte ebenso hoch signifikante Werte für die Unterschiede zwischen den beiden Variantengruppen ($p = 0,0015$ bzw. $p = 0,001$).

Dass auch die Streuungen der beiden Gruppen sich deutlich unterscheiden, (Standardabweichungen: 0,64 für ‚falsche‘ Varianten und 1,21 für ‚richtige‘ Varianten der Referenzmelodie) ist nicht verwunderlich, denn es ist plausibel anzunehmen, dass die Versuchspersonen die Ähnlichkeit der Melodien, die eine Verwandtschaft mit der Referenzmelodie aufwiesen, differenzierter beurteilten als die Melodien, die keinen Zusammenhang mit der Referenzmelodie besaßen. Der Levene-Test auf Varianzgleichheit nähert sich entsprechend mit $p = 0,53$ einem signifikanten Wert an.

Die inhaltliche Hypothese 1 kann also bestätigt werden: Varianten mit einem anderen Ursprung als der zum Vergleich angebotenen Referenzmelodie werden signifikant unähnlicher zur Referenzmelodie beurteilt als Melodien, deren Ursprung die Referenzmelodie ist.

4.3.2.3.3 Zusammenhang der Ähnlichkeitsbewertungen in Experiment 1 und im Kontrollexperiment

Die Hypothesen 2, 3a und 3b wurden in einem dreistufigen Vorgehen überprüft:

1. Zunächst sollte ein t-Test für gepaarte Stichproben prüfen, ob die Beurteilungswerte der in beiden Experimenten bewerteten Varianten im Kontrollexperiment im Mittel von den entsprechenden Bewertungen aus Experiment 1 signifikant abweichen. Dabei ist die Richtung des Effekts nicht eindeutig vorhersagbar. Denn sowohl eine Verschiebung in Richtung der maximalen Ähnlichkeit ist durch den Kontext der Varianten mit fremdem Ursprung denkbar, als auch ein Effekt der Transpositionen im Kontrollexperiment, der sich in Richtung größerer Unähnlichkeit manifestieren könnte, scheint plausibel. Ein t-Test mit zweiseitiger Signifikanzprüfung scheint also geboten zu sein.
2. Danach soll mit der Korrelation nach Pearson-Bravais geprüft werden, wie stark der Zusammenhang der Ähnlichkeitsurteile bei den selben Varianten in beiden Experimenten ist.
3. Schließlich kann eine Rangkorrelation nach Spearman noch darüber Auskunft geben, ob zusätzlich zum absoluten Zusammenhang auch ein Zusammenhang in Hinsicht der Rangreihenfolge der bewerteten Varianten besteht. Da ein positiver Zusammenhang zu erwarten ist, kann eine einseitige Signifikanzprüfung vorgenommen werden.

In Experiment 1 und im Kontrollexperiment waren die Melodien A und C mit ihren sechs zugehörigen Varianten getestet worden. Die Mittelwerte der Bewertungen aus Experiment 1 und dem Kontrollexperiment unterschieden sich um 0,98 zugunsten des Kontrollexperiments, wie die folgende Tabelle zeigt:

Statistik bei gepaarten Stichproben

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	KÄHNAC	4,6183	12	1,13216	,32683
	EX1ÄHNAC	3,6389	12	1,51415	,43710

Tabelle 4-18: Mittlere Ähnlichkeitsbewertungen der Melodie-Variantenpaare zu den Referenzmelodien A und C in Experiment (KÄHNAC) und in Experiment 1 (EX1ÄHNAC).

Auch die Standardabweichungen zeigen einen leichten Unterschied von 0,38 Einheiten. Der t-Test für abhängige Stichproben ergab folgerichtig eine signifikante Irrtumswahrscheinlichkeit für den errechneten t-Wert:

Test bei gepaarten Stichproben

		Gepaarte Differenzen				T	df	Sig. (2-seitig)	
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere				Obere
Paaren 1	KÄHNAC - EX1ÄHNAC	,9794	,78572	,22682	,4802	1,4787	4,318	11	,001

Tabelle 4-19: t-Test für gepaarte Stichproben für die Mittelwertunterschiede zwischen den Ähnlichkeitsbewertungen der Melodie-Variantenpaare der Referenzmelodie A und C in Experiment (KÄHNAC) und in Experiment 1 (EX1ÄHNAC).

Die inhaltliche Hypothese 2 (bzw. ihre entsprechende Null-Hypothese) muss also abgelehnt werden: Es besteht ein signifikanter Unterschied zwischen den Bewertungen der selben Varianten in Experiment 1 und im Kontrollexperiment. Im Kontext der sehr verschiedenen Varianten, die keine Verwandtschaft mit der Referenzmelodie besaßen, wurden die mit der Referenzmelodie verwandten Varianten ähnlicher beurteilt als ohne diesen Kontext. Die Beurteilungen sind im Kontrollexperiment im Mittel um 0,98 Einheiten zum Pol der maximalen Ähnlichkeit hin verschoben.

Ein Effekt der Transposition im Kontrollexperiment ist in den Daten nicht zu erkennen. Wenn die Transposition überhaupt eine Rolle bei der Ähnlichkeitsbewertung gespielt hat, dann ist ihr Effekt zumindest kleiner als der des Kontextes der übrigen Vergleichsmelodien.

Für den Zusammenhang zwischen den Ähnlichkeitsurteilen ergaben sich hohe und hoch signifikante Werte:

r (Pearson) = 0,862; $p < 0,000$

ρ (Spearman) = 0,714; $p = 0,005$

Beide Werte stützen die Hypothesen 3a und 3b. Es besteht also ein signifikanter Zusammenhang zwischen den Bewertungen der Varianten in Experiment 1 und im Kontrollexperiment, und zwar sowohl in Hinsicht der absoluten Differenzen der Werte von ihrem Mittelwert (Pearson) als auch hinsichtlich der Rangplätze der Ähnlichkeitswerte (Spearman).

Insgesamt ist festzuhalten, dass ein starker Zusammenhang zwischen den Ähnlichkeitsurteilen von Experiment 1 und denen des Kontrollexperimentes besteht. Dieser starke Zusammenhang ergab sich trotz der Tatsache, dass es sich bei den Versuchspersonen in Haupt- und Kontrollexperiment zum größten Teil um unterschiedliche Personen handelte und trotz des verschiedenen Beurteilungskontextes in den unterschiedlichen Experimentaldesigns. Die Hinzunahme von stark verschiedenen Melodievarianten in den Beurteilungskontext ‚verschob‘ die Ähnlichkeitsurteile für die identischen Varianten jedoch in Richtung größerer Ähnlichkeit.

In jedem Fall konnten die Musikexperten die Varianten anderer Melodien von den Varianten der Referenzmelodie trotz der vorgenommenen Transpositionen unterscheiden. Es wurde keine einzige ‚falsche‘ Variante ähnlicher beurteilt als eine ‚richtige‘. Die Ähnlichkeitsmittel der ‚falschen‘ und der ‚richtigen‘ Varianten unterschieden sich hoch signifikant.

4.3.3 Experiment 3 (Das Kontrollexperiment 2)

In Experiment 1 waren bewusst bei der Konstruktion der Varianten keine Intervallfehler mit Konturänderung und solche ohne Konturänderung kombiniert worden, um den Effekt von Konturänderungen auf die Ähnlichkeitsbeurteilungen erheben zu können. Die Einflüsse von Konturfehlern und ‚reinen‘ Intervallfehlern hatten sich als gleichbedeutend für die Ähnlichkeitsurteile der Experten erwiesen. In den entsprechenden statistischen Tests waren keine Unterschiede festzustellen gewesen(s.o.).

Realistischerweise finden sich aber in Melodievarianten – seien es schlecht wiedergegebene Melodien aus dem Gedächtnis oder Varianten eines Volksliedes – in der Regel sowohl Konturabweichungen als auch Intervalländerungen. Im zweiten Kontrollexperiment sollte geprüft werden, ob die getesteten Ähnlichkeitsalgorithmen sich auch in diesem Fall

bewähren. Abgesehen von der Hypothese, dass das optimale Ähnlichkeitsmaß für Experiment 1 auch hinsichtlich dieser neuen, unter differierenden experimentellen Bedingungen gewonnenen Daten, eine gute Erklärungsleistung erbringt, wurden keine speziellen Hypothesen für Experiment 3 aufgestellt.

4.3.3.1 Methode

Versuchspersonen:

Teilnehmer des Experimentes waren 10 Studierende und Absolventen der Musikwissenschaft. Wie bei den anderen Experimenten zeichneten sie sich durch einen sehr aktiven musikalischen Hintergrund aus.

Versuchsmaterialien:

Aus den Referenzmelodien des Experiments 1 wurden vier ausgewählt, zu denen jeweils 8 Varianten konstruiert wurden. Alle Varianten wurden wie in Experiment 1 und im Gegensatz zu Experiment 2 von der Referenzmelodie ausgehend konstruiert.

Die Fehlerwahrscheinlichkeiten waren gleich zu denen aus Experiment 1, mit der Ausnahme, dass Intervallfehler mit und ohne Konturänderung zu einer Fehlerklasse mit $p=0,6$ zusammengefasst wurden. Alle möglichen Kombinationen der verschiedenen Ausprägungen von Intervall- und Konturfehlern (0,3,6,9 Fehler jeweils bei Intervallen und Kontur) wurden auf die insgesamt 21 Varianten mit Intervallfehlern gleichmäßig verteilt. Insgesamt nahm also die mittlere Anzahl der Intervallfehler pro Variante zu und lag mit durchschnittlich 10 Fehlern (Bereich: 0 bis 25) deutlich über der von Experiment 1.

Alle Varianten wurden den Versuchsteilnehmern transponiert gegenüber der Referenzmelodie dargeboten. Die Auswahl der Transpositionstufe für eine einzelne Melodie folgte den selben Regeln wie in Experiment 2 (s.o.).

Prozedur:

Es wurde die gleiche Versuchsprozedur verwendet wie in den Experimenten 1 und 2. Die Versuchspersonen hatten 4x8 Melodiepaare auf ihre Ähnlichkeit zu bewerten, in diesem Experiment allerdings auf einer Skala von 1-10 (10=maximale Ähnlichkeit), um den Versuchspersonen einen größeren Spielraum zur Differenzierung zu erlauben und gleichzeitig die Gültigkeit des linearen Modells auch bei einer anderen Skalierung der abhängigen Va-

riable zu überprüfen. Nach den *Ähnlichkeitsratings* füllten die Versuchsteilnehmer den gleichen Fragebogen zu ihrem musikalischen Hintergrund aus, wie die Teilnehmer der beiden vorangegangenen Experimente.

Design:

Die Versuchspersonen wurden in nur einer Gruppensession getestet. Zur Überprüfung der Urteilsstabilität wurden zwei Vergleichspaare im Verlauf der Testsession wiederholt. Die Differenz zwischen den Bewertungen der gleichen Vergleichspaare durfte im Durchschnitt nicht mehr als 1 betragen. Anders formuliert: Es durfte bei einem Paar nur eine Abweichung von 2 auftreten, wenn bei dem anderen gleichzeitig eine Abweichung von 0 gegeben war.

Um die Reliabilität zu prüfen, waren insgesamt drei – bis auf die Transposition – identische Varianten in den 32 Vergleichspaaren enthalten. Von den drei identischen Varianten musste eine Vpn. mindestens zwei mit einem Ähnlichkeitswert von mindestens 9 bewerten, damit ihre Daten in die Auswertungen eingingen.

Rechnerische Behandlung der Daten:

Die erhobenen Daten der zehn Versuchspersonen wurden nach den folgenden Schritten selektiert:

- Es wurden alle Vpn. ausgeschlossen, die nicht mindestens zwei identische Varianten mit mindestens dem Wert 9 beurteilt hatten. Es blieben alle 8 Vpn. in der Auswertung.
- Es wurden alle Vpn. ausgeschlossen, deren Beurteilungen der 2x2 identischen Melodie-Variantenpaaren eine durchschnittliche Differenz von mehr als 1 auswies. Es blieben 5 Musikexperten in der Auswertung.

Da sich in Experiment 1 keine auch nur annähernd signifikanten Lern- und Reihenfolgeeffekte gezeigt hatten, wurde auf die aufwendige Überprüfung dieser Effekte in den Daten des Kontroll-experiments verzichtet.

4.3.3.2 Ergebnisse

Ein weiteres Mal zeigte sich eine große Übereinstimmung zwischen den Urteilen der selektierten Musikexperten. Cronbach's α erreichte einen Wert von 0,948 über die Bewer-

tungen der 32 items der fünf Experten. Auch die anderen Maße, mit denen sich der Zusammenhang zwischen den Urteilen der fünf Experten messen lässt, wiesen entsprechend hohe Werte aus: Kaiser-Meyer-Olkin: 0,851; *Measure of Sampling Adequacy*: keiner der fünf Werte unter 0,77; Bartlett-Test auf Sphärizität: $p < 0,001$ für die Annahme, dass keine Korrelation zwischen den Variablen (Expertenurteilen) in der Grundgesamtheit besteht. Auch unter diesen gegenüber Experiment 1 deutlich veränderten Versuchsbedingungen – andere Art der Variantenkonstruktion, andere Fehlerverteilung, Transposition der Vergleichsvarianten, Einschätzung auf zehnstufiger statt auf siebenstufiger Skala – zeigen die Urteile der Musikexperten einen extrem hohen Zusammenhang. Die Vorstellung davon, was musikalische Ähnlichkeit ist und wie sie auf *Ratingskalen* zu beurteilen ist, scheint zumindest unter Musikexperten stabil und valide zu sein.

Die musikalischen Aktivitäten der selektierten Versuchspersonen dieses Kontrollexperiments waren ähnlich stark ausgeprägt wie in den beiden vorangegangenen Experimenten: Im Mittel spielten die fünf Experten seit 16,2 Jahren ihre Instrumente, hatten 76 Monate bezahlten Unterricht genossen und in ihrer aktivsten Phase 15,4 Stunden pro Woche Musik gemacht.

4.4 Modellierung menschlicher Urteile mit Ähnlichkeitsalgorithmen

4.4.1 Ermittlung des optimalen Ähnlichkeitsmaßes für das Hauptexperiment (Lineare Regression)

Primäres Ziel der in diesem Kapitel beschriebenen Untersuchung war die Ermittlung eines algorithmischen Ähnlichkeitsmaßes, welches über alle getesteten Varianten hinweg möglichst nahe an dem Urteil der selektierten Musikexperten liegen sollte. Dabei liegt es nahe zu vermuten, dass das optimale Ähnlichkeitsmaß nicht ein einzelnes algorithmisches Ähnlichkeitsmaß darstellt, sondern eine Kombination von Maßen, die verschiedene musikalische Dimensionen der Testmelodien einbeziehen, der menschlichen Wahrnehmung näher kommt. Ähnlich beschreibt auch Selfridge-Field (1998), dass jede Kombination von zwei Ähnlichkeitsmaßen bessere Ergebnisse liefert als ein einzelnes Maß.

Die oben ausführlich dargestellten Maße decken fünf wahrnehmungsbezogene Dimensionen ab, die alle für die Ähnlichkeitsurteile von Musikexperten eine Rolle spielen können. Diese fünf Dimensionen sind: Tonhöhen- bzw. Intervallstruktur, Tonhöhenkontur, Rhyth-

mus, implizierte Harmonik und kurze (charakteristische) Motive. Die oben vorgestellten Maße lassen sich diesen fünf Dimensionen auf folgende Weise zuordnen:

Dimension	Intervalle	Kontur	Rhythmus	Implizite Harmonik	Kurze Motive
Definition	Maße, die auf der Reihe der Tonhöhen, der Intervalle oder deren klassifizierten Werten aufbauen	Maße, die auf einer Konturersatzreihe (Interpolations- oder Rangwerte) aufbauen	Maße, die auf der Sequenz der klassifizierten Dauernwerte oder der gaußifizierten Onsetwerte aufbauen	Maße, die auf der Sequenz der harmonisch gewichteten Tonhöhenwerte aufbauen	Maße, die auf kurzen Untersequenzen von Tonhöhenwerten, klassifizierten Intervallen oder Intervallrichtungen aufbauen
Maße	diff, diffed, rawed, rawedw, rawpc, rawpcst, rawpcw, rawpcwst	Consed, conspc, conspcst, coned, conpc, conpcst, fourr, fourrst, fourrw, fourrwst, ngrsumcr, ngrukkor, ngrcoorr	Rhythfuzz, rhythgaus, ngrcoorfr, ngrsumfr, ngrukkfr	harmcore	ngrsumco, ngrukkon, ngrcoord, ngrsumcf, ngrukkof, ngrcoorf

Tabelle 4-20: Zuordnung der verwendeten Ähnlichkeitsmaße zu fünf inhaltlichen Dimensionen.

Zum Auffinden des optimalen Maßes bzw. der optimalen Kombination von Maßen, die am besten mit den *ratings* der Experten übereinstimmt, wurde folgender, zweistufiger Algorithmus angewendet.

- Summiere über alle 84 Varianten des Hauptexperimentes die euklidische Distanz jedes Maßes zum Mittelwert der *ratings* der Versuchspersonen nach der Formel

$$d(x, y) = \sqrt{\sum_{i=1}^{84} (Y_i - X_i)^2} \quad \text{mit } Y_i \text{ als Mittelwert der } \textit{Expertenratings} \text{ für}$$

Melodievariante i und X_i als Wert eines algorithmischen Ähnlichkeitsmaßes für Melodievariante i . Ermittle für jede der fünf Dimensionen das Maß, das den kleinsten Distanzwert hat. Diese Maße sind die besten Maße für die fünf Dimensionen.

- Rechne eine lineare Regression mit diesen besten Maßen. Ermittle die beste Regression mit diesen Maßen. Als Kriterien für die Evaluation der Regressionslösung dienen:
 - das Vorzeichen eines Regressionskoeffizienten muss positiv und

- er muss auf einem Niveau von $p < 0,05$ signifikant sein,
- die Summe der erklärten quadrierten Abweichungen muss die Summe der residualen quadrierten Abweichungen deutlich übersteigen, sodass ein F-Test einen p-Wert von $< 0,05$ erreicht,
- das korrigierte R^2 soll maximal sein und
- der Standardfehler soll minimal sein.

Dieses Verfahren wurde sowohl für die Auswertung des Hauptexperiments als auch der Kontrollexperimente vorgenommen.

Die Zuordnung der getesteten Maße zu fünf verschiedenen inhaltlichen Dimensionen ist auch in Blick auf das hier verwendete Optimierungsverfahren der linearen Regression sinnvoll. Denn eine der Forderungen an die Faktoren eines Regressionsmodells ist, dass sie weitestgehend voneinander unabhängig sein sollen, um Kollinearitäts- und Suppressionseffekte zu vermeiden. Die wird durch die Auswahl nur je eines Faktors pro inhaltlicher Dimension im Optimierungsalgorithmus gewährleistet. Maße, die vermutlich hoch miteinander korrelieren, wie z.B. *coned* und *conpc* oder *rhythfuzz* und *rhythgaus*, können nicht gleichzeitig für ein Regressionsmodell ausgewählt werden. Diese Beschränkung auf fünf potentielle Regressionsfaktoren vereinfacht auch deutlich das Auffinden der optimalen Faktorenkombination.

Für das Hauptexperiment ergaben sich euklidischen Distanzen für die einzelnen Ähnlichkeitsmaße, wie sie in der folgenden dreispaltigen Tabelle dargestellt sind. Die Tabelle ist von links oben nach rechts unten zu lesen. Die geringste Distanz von 0 hat natürlich das Mittel der *Expertenratings* zu sich selbst (*vpn_mean*), gefolgt von der Edit Distance, die auf der Konturersatzreihe von Müllensiefen-Frieler aufbaut (*coned*; Distanz 5,29) und der gewichteten Edit Distance der Tonhöhenreihen (*rawedw*; Distanz 5,63). Auf den letzten Plätzen findet sich neben allen Rhythmusmaßen das Modell von Selfridge-Field. Die Spannweite der euklidischen Distanzen der algorithmischen Maße reicht von 5,29 bis 11,18. Zur Kontrolle wurde noch die Distanz der Versuchspersonen-Urteile zu einer Konstante, nämlich 0,5, mit in die Tabelle aufgenommen (0,5 *KONST*). Diese zeigt mit Abstand die größte Distanz, und das trotz der Tatsache, dass die Urteile der Versuchspersonen im Mittel 4,0 betragen, also genau auf der Hälfte der Ähnlichkeitsskala liegen. Insofern kann man allen hier getesteten Ähnlichkeitsmaßen zugute halten, dass sie weit differen-

ziertere und bessere Ähnlichkeitsschätzungen liefern als der Mittelwert der Expertenurteile.

VPN_MEAN	0	CONSED	7,29513447	CONSPC	9,99224417
CONED	5,29245469	CONPCST	7,81920546	FOURRI	10,0102487
RAWEDW	5,63380719	CONPC	8,00339203	FOURRST	10,1054413
RAWED	5,80486154	RAWPCST	9,18246857	NGRUKKOR	10,1325741
NGRCOORD	5,9404196	RAWPC	9,27255174	DIFF	10,1738995
NGRUKKON	6,08484904	RAWPCWST	9,4813044	NGRSUMCR	10,3803696
NGRSUMCO	6,16556935	RAWPCW	9,62004454	RHYTFUZZ	10,4311813
HARMCORE	6,17690398	FOURRW	9,68105158	ESFMODEK	10,6393058
DIFFED	6,30159649	FOURRWST	9,72024592	RHYTGAUS	10,6954894
NGRUKKOF	6,68471942	FOURR	9,78068977	NGRUKKFR	11,0532836
NGRCOORDF	6,71865164	CONSPCST	9,84445712	NGRCOORDFR	11,1158631
NGRSUMCF	7,21157162	NGRCOORDR	9,92365661	NGRSUMFR	11,1832575
				0,5 KONST	21,89091372

Tabelle 4-21: Euklidische Distanzen der verwendeten Ähnlichkeitsmaße zum Mittel der Urteile der selektierten Experten über die 84 Melodie-Variantenpaare aus Experiment 1.

Für die fünf Dimensionen wurden als beste Maße ermittelt:

- Tonhöhen/Intervalle: *rawedw* (5,63)
- Kontur: *coned* (5,29)
- Rhythmus: *rhythfuzz* (10,43)
- Implizierte Harmonik: *harmcore* (6,18)
- Motive: *ngrcoord* (5,94)

Diese besten Maße wurden als Regressoren für eine Serie von linearen Regressionen verwendet. Dabei wurde der in SPSS implementierte Algorithmus „Schrittweise“ als Ausgangspunkt benutzt, der die einzelnen Regressionsfaktoren nach dem Betrag ihres β -Gewichts in die Modelle schrittweise aufnimmt und entfernt und dabei die Zunahme des korrigierten R^2 prüft. Über diesen Algorithmus hinaus wurden noch weitere Regressionsmodelle mit anderen Kombinationen von Regressoren ‚per Hand eingegeben‘ und gerechnet, da das „Schrittweise“-Verfahren nicht zwingend die optimale Kombination von Regressoren finden muss (vgl. das Beispiel von Thompson (1995), das Bortz (1999, S. 447) anführt). Im Falle des Hauptexperiments gelangte das SPSS-Verfahren, wie es in folgenden Tabellen dokumentiert ist, jedoch zur optimalen Regressorenkombination:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,831 ^a	,691	,687	,8866175
2	,867 ^b	,751	,745	,8000178
3	,911 ^c	,831	,824	,6642181
4	,911 ^d	,830	,826	,6618007

a. Einflußvariablen : (Konstante), CONED

b. Einflußvariablen : (Konstante), CONED, NGRCOORD

c. Einflußvariablen : (Konstante), CONED, NGRCOORD, RAWEDW

d. Einflußvariablen : (Konstante), NGRCOORD, RAWEDW

ANOVA^e

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	144,148	1	144,148	183,373	,000 ^a
	Residuen	64,459	82	,786		
	Gesamt	208,607	83			
2	Regression	156,765	2	78,382	122,467	,000 ^b
	Residuen	51,842	81	,640		
	Gesamt	208,607	83			
3	Regression	173,312	3	57,771	130,944	,000 ^c
	Residuen	35,295	80	,441		
	Gesamt	208,607	83			
4	Regression	173,131	2	86,565	197,647	,000 ^d
	Residuen	35,476	81	,438		
	Gesamt	208,607	83			

a. Einflußvariablen : (Konstante), CONED

b. Einflußvariablen : (Konstante), CONED, NGRCOORD

c. Einflußvariablen : (Konstante), CONED, NGRCOORD, RAWEDW

d. Einflußvariablen : (Konstante), NGRCOORD, RAWEDW

e. Abhängige Variable: VPN_MEAN

Koeffizienten^f

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	2,323	,157		14,779	,000
	CONED	3,758	,278	,831	13,542	,000
2	(Konstante)	1,668	,205		8,150	,000
	CONED	2,488	,380	,550	6,546	,000
	NGRCOORD	2,144	,483	,373	4,440	,000
3	(Konstante)	,219	,291		,753	,454
	CONED	,305	,476	,068	,641	,523
	NGRCOORD	2,682	,410	,467	6,534	,000
	RAWEDW	3,111	,508	,503	6,124	,000
4	(Konstante)	,092	,213		,434	,666
	NGRCOORD	2,852	,312	,497	9,150	,000
	RAWEDW	3,355	,336	,543	9,999	,000

a. Abhängige Variable: VPN_MEAN

Tabellen 4-22a, 4-22b und 4-22c: Gütewerte und Koeffizienten der Regressionsmodelle, die durch das *Schrittweise*-Verfahren ermittelt werden (Regression der Vpn-Urteile auf die besten ausgewählten Ähnlichkeitsmaße).

Die optimale Kombination zur Vorhersage der *Expertenratings* bilden die Variablen *rawedw* und *ngrcoord*. Interessanterweise konnte das Maß *coned*, welches den geringsten Abstand aller Maße zu den *Expertenratings* aufwies, in Kombination mit *rawedw* keinen hohen explikativen Beitrag mehr leisten. Während alle übrigen Kriterien auch für andere Modelle zutrafen, stellen sich bei dieser Regressorenkombination das größte korrigierte R^2 (0,826) und der kleinste Standardfehler der Schätzung (0,662) ein. Mit 83% liegt die erklärte Varianz (R^2) dieses Modells recht hoch, sodass man von einem aussagekräftigen Modell sprechen kann.

Da das konstante Glied mit einem Betrag von 0,092 aufgrund seiner hohen Irrtumswahrscheinlichkeit nicht in dieses Modell aufgenommen wird, lautet die resultierende Regressionsgleichung für das optimale Modell zur Vorhersage der *Expertenratings* auf der 7-stufigen Skala:

$$\hat{y} = 3,355 \cdot \text{rawedw} + 2,852 \cdot \text{ngrcoord}$$

Mit Hilfe dieser Regressionsgleichung wurde die neue Variable *optil* gebildet und deren euklidischer Abstand zu den *Expertenratings* bestimmt: Mit 3,789 ist dieser Abstand um 28,5% kleiner als der des besten Einzelmaßes *coned* (5,29). Der Optimierungsprozeß der Ähnlichkeitsmaße über die lineare Regression war demnach erfolgreich, wie auch das folgende Diagramm zeigt.

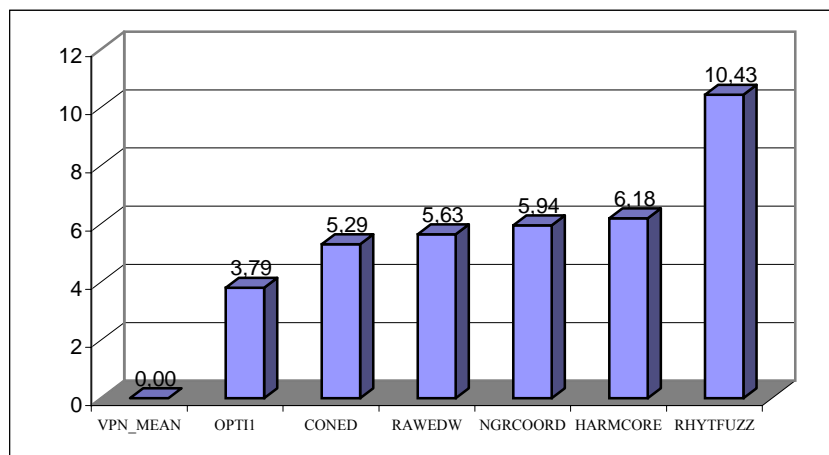


Abb. 4-14: euklidische Distanz der besten Ähnlichkeitsalgorithmen im Vergleich mit den mittleren *Ähnlichkeitsratings* der Vpn. über alle 84 Varianten von Exp.1

In der Normierung auf den Bereich von 0 bis 1 ergaben sich die folgenden β -Gewichte für die Regressionsgerade von *optil*:

$$\hat{y} = -0,151 + 0,559 \cdot \text{rawedw} + 0,475 \cdot \text{ngrcoord}$$

(Zu beachten ist, dass das Maß formal auch Werte <0 ergeben kann. Dieser Fall spielt jedoch in der Praxis keine Rolle)

4.4.2 Ermittlung der optimalen Ähnlichkeitsmaße für das 1. Kontrollexperiment (Exp. 2)

Ganz analog zum im vorigen Absatz beschriebenen Vorgehen wurden die optimalen Ähnlichkeitsmaße für die Daten des Kontrollexperiments ermittelt: Die besten Einzelmaße der fünf Dimensionen wurden mit Hilfe der euklidischen Distanzen bestimmt und eine lineare Regression verwendete diese fünf besten Einzelmaße als Regressoren, von denen die mit einem relevanten Erklärungsbeitrag ausgewählt und untereinander gewichtet wurden. Da das Vorgehen identisch zu dem im Hauptexperiment ist, wird im folgenden auf eine ausführliche Dokumentation mit allen Einzelergebnissen und -tabellen verzichtet, sodass nur die relevanten Ergebnisse im Text referiert werden. Die zugehörigen Tabellen finden sich im Anhang.

Die Gesamtmenge der Melodievarianten des Kontrollexperiments wurden jedoch zunächst zweigeteilt in die ‚richtigen‘ Varianten, die von der jeweiligen Referenzmelodie abstammten und sich durch mehr oder minder starke Abweichungen von ihr unterschieden, und in jene ‚falschen‘ Varianten, die ihren Ursprung in einer anderen Referenzmelodie hatten und keine (offensichtliche) Ähnlichkeit mit der jeweiligen Referenzmelodie besaßen. Zunächst wurde ein optimales Ähnlichkeitsmaß nur für die richtigen Varianten ermittelt, welches mit dem optimalen Ähnlichkeitsmaß des Hauptexperiments zu vergleichen ist. Danach wurde das optimale Ähnlichkeitsmaß für die vereinigte Menge der ‚richtigen‘ und ‚falschen‘ Varianten berechnet.

4.4.2.1 Die ‚richtigen‘ Varianten:

Das Kontrollexperiment beinhaltete insgesamt 13 ‚richtige‘ Melodievarianten. Die euklidischen Abstände der getesteten Maße vom Mittel der *ratings* der Versuchspersonen gibt die folgende Tabelle wieder.

VPN_MEAN	0	CONSED	2,10541776	FOURRWST	3,03923815
DIFFED	1,29688675	FOURRI	2,20840289	ESFMODEK	3,0556805
RAWED	1,59301343	CONSPCST	2,3319825	NGRCOOFr	3,08975553
RAWEDW	1,77116553	CONED	2,34609514	RAWPC	3,17147708
NGRSUMCO	1,88098309	CONSPC	2,38196822	RAWPCWST	3,21896461
NGRSUMCF	1,90031402	NGRUKKOR	2,50009986	RAWPCW	3,36731624
NGRCOORF	1,96472735	NGRCOORR	2,56138892	NGRUKKFR	3,41834328
HARMCORE	1,98429143	NGRSUMCR	2,57443529	0,5 KONST	3,46410161
NGRUKKON	1,98617201	FOURR	2,8958865	NGRSUMFR	3,50605419
DIFF	1,99107607	FOURRST	2,8958865	CONPCST	3,81939377
NGRCOORD	2,01933894	RAWPCST	3,01501694	CONPC	3,86445532
NGRUKKOF	2,02451375	FOURRW	3,03923815	RHYTFUZZ	3,88172081
				RHYTGAUS	4,31973804

Tabelle 4-23: Euklidische Distanzen der algorithmischen Ähnlichkeitsmaße zu den *ratings* der Vpn. über die 13 richtigen Varianten in Exp. 2.

Als beste Maße mit der geringsten Distanz zu den Urteilen der Versuchspersonen erwiesen sich *diffed* (Distanz zu Vpn-Urteilen: 1,3), *ngrsumco* (1,88), *harmcore* (1,98), *consed* (2,11), *ngrcoorfr* (3,09). Die Distanzen reichten insgesamt von 1,3 bis 4,3.

Das beste Regressionsmodell enthielt die zwei Maße *ngrsumco* und *harmcore*. Es ergaben sich mit diesen beiden Regressoren sehr hohe Werte für die Modellgüte: $R = 0,960$, $R^2 = 0,922$, korrigiertes $R^2 = 0,906$ und ein Standardfehler der geschätzten Werte von 0,37. Es werden also 92% der Streuung in den *Ratingdaten* der Musikexperten durch das Modell erklärt, und die mittlere Abweichung der geschätzten von den empirischen Werte beträgt 0,37 Einheiten auf der 7-stufigen Skala. Zu den Koeffizienten der beiden Regressionsfaktoren kommt noch ein konstantes Glied mit einem Wert von 2,294 hinzu. Diese konstante Versetzung in Richtung des Pols der maximalen Ähnlichkeit stimmt mit der oben referierten Verschiebung der Ähnlichkeitswerte überein, die auf den Kontext der Stimuli des Experiments zurückzuführen ist. In der Umgebung von ‚falschen‘ und ‚richtigen‘ Varianten werden die ‚richtigen‘ Melodievarianten ähnlicher beurteilt als ohne diesen Kontext der ‚falschen‘ Varianten.

Ngrsumco und *harmcore* sind jedoch nicht die beiden Faktoren, aus denen das optimale Ähnlichkeitsmaß des Hauptexperimentes gebildet wurde, obwohl die ‚richtigen‘ Varianten des Kontrollexperimentes eine echte Teilmenge der Melodievarianten des Hauptexperimentes darstellen. Um die Validität des optimalen Maßes des Hauptexperimentes zu prüfen, wurden die Daten der 13 ‚richtigen‘ Varianten ein zweites Mal einer Regressionsanalyse

unterzogen, dieses Mal jedoch mit den fünf besten Maßen des Hauptexperiments. Wieder bildeten *rawedw* und *ngrcoord* das beste Regressionsmodell, ergänzt dieses Mal durch ein konstantes Glied von 2,254, welches auf den Kontext der Stimuli zurückzuführen ist. Dieses Modell, welches im folgenden Modell 1 genannt werden soll, zeichnete sich ebenso durch hohe Werte der Modellgüte aus: $R = 0,946$, $R^2 = 0,895$, korrigiertes $R^2 = 0,874$, Standardfehler der geschätzten Werte = 0,43.

Danach wurde schließlich noch genau der entgegengesetzte Fall durchgerechnet: Regressionsanalysen mit den besten Maßen des Kontrollexperiments (Modell 2) auf den Daten des Hauptexperiments. Wieder bestand die optimale Kombination aus *ngrsumco* und *harmcore*. Das Modell erreichte folgende Werte der Anpassungsgüte: $R = 0,862$, $R^2 = 0,744$, korrigiertes $R^2 = 0,737$ und Standardfehler der geschätzten Werte = 0,81.

Die standardisierten β -Gewichte waren für beide Modelle auf beiden Datensätzen in etwa die gleichen. *Rawedw* wurde etwa 1,15mal stärker gewichtet als *ngrcoord* und *ngrsumco* ungefähr 1,05mal stärker als *harmcore*.

Es liefern also beide Paare von Maßen gute Schätzungen der Urteile der Versuchspersonen in der Regressionsanalyse. Dies ist auch nicht verwunderlich, wenn man die Tabellen der euklidischen Abstände betrachtet. Dort finden sich in der Tabelle des Hauptexperiments *ngrsumco* und *harmcore* sehr dicht hinter *rawedw* und *ngrcoord* und umgekehrt finden sich in der Tabelle zu den ‚richtigen‘ Varianten des Kontrollexperiments *rawedw* und *ngrcoord* nicht weit entfernt von *ngrsumco* und *harmcore*. Die Maße unterschieden sich insgesamt also nicht besonders stark in ihren euklidischen Distanzen zu den Bewertungen der Versuchspersonen.

Dennoch sprechen zwei Gründe dafür, das Modell mit den Maßen *rawedw* und *ngrcoord* (Modell 1) als das insgesamt überlegene anzusehen: Zum einen erklärt dieses Modell einen größeren Datensatz mit Daten von 84 Melodievarianten, die von 23 Versuchspersonen erhoben wurden. Zum zweiten fällt der Unterschied der korrigierten R^2 -Werte zwischen den beiden Modellen für die beiden Datensätze zugunsten des Modells mit *rawedw* und *ngrcoord* (Modell 1) aus:

- Differenzen der korrigierten R^2 –Werte für das Kontrollexperiment: $0,906 - 0,874 = 0,032$ zugunsten Modell 2;
- Differenzen der korrigierten R^2 –Werte für das Hauptexperiment: $0,781 - 0,737 = 0,044$ zugunsten Modell 1;

Aus diesen Gründen wird die oben gefundene Linearkombination aus *rawedw* und *ngrcoord* insgesamt für die Schätzung der Ähnlichkeitswerte von Musikexperten für ein Experiment, in dem einer Referenzmelodie echte Varianten zur Ähnlichkeitsbeurteilung gegenübergestellt werden, als das geeignete Modell angenommen.

Freilich unterscheiden sich die Gleichungen anhand der Koeffizienten und des konstanten Gliedes. Für eine Schätzung der Ähnlichkeiten in einer Aufgabe, bei der nur Varianten einer Melodie auf ihre Ähnlichkeit mit der Referenzmelodie verglichen werden, ist

$$\hat{y} = 3,355 \cdot \text{rawedw} + 2,852 \cdot \text{ngrcoord}$$

die beste Schätzung. Bei Beurteilungsaufgaben, bei denen ‚falsche‘ und ‚richtige‘ Varianten den Beurteilungskontext bilden, sind die Schätzungen für die Ähnlichkeit der richtigen Varianten am besten durch das kombinierte Maß (*opti2*) mit der Formel

$$\hat{y} = 2,254 + 2,61 \cdot \text{rawedw} + 1,72 \cdot \text{ngrcoord}$$

bzw.

$$\hat{y} = 0,209 + 0,435 \cdot \text{rawedw} + 0,287 \cdot \text{ngrcoord}$$

für die Normierung auf den Bereich von 0 bis 1 zu beschreiben. Die euklidische Distanz zwischen diesem kombinierten Maß und den Urteilen der Versuchspersonen beträgt über die 13 richtigen Varianten gerechnet 1,137 und ist damit 12,3% besser als das beste Einzelmaß *diffed* in diesem Vergleich.

4.4.2.2 Die ‚richtigen‘ und ‚falschen‘ Varianten:

Zwei Fragen drängen sich hinsichtlich der optimalen Maße für das gesamte Kontrollexperiment auf:

Welches Maß hat die geringste Distanz zu den Urteilen der Musikexperten für alle 24 Varianten?

Welches Maß unterscheidet so wie die Versuchspersonen (s.o.) immer korrekt zwischen ‚falschen‘ und ‚richtigen‘ Varianten, sodass ein Ähnlichkeitswert als Trennungsmarke für falsche und richtige Varianten genutzt werden kann?

Die Tabelle mit den euklidischen Distanzen der getesteten Maße zu den Urteilen der Versuchspersonen dient als Ausgangsbasis für die Ermittlung des optimalen Maßes für alle Varianten des Kontrolltests (Frage 1).

VPN_MEAN	0	NGRSUMCR	3,30150345	RAWPCW	5,20686753
DIFFED	2,04258105	NGRCOORR	3,330593	FOURRWST	5,3029391
RAWED	2,19668884	CONSED	3,57206245	FOURR	5,45782753
NGRUKKON	2,43956243	CONED	3,6295554	FOURRST	5,45782753
NGRCOORD	2,50546508	RHYTFUZZ	3,64994731	RAWPCST	5,49469773
NGRSUMCO	2,51334925	RHYTGAUS	3,80593719	CONPCST	5,55753837
NGRCOORF	2,58811776	NGRUKKFR	3,91766035	RAWPC	5,73617046
NGRUKKOF	2,59906584	NGRSUMFR	4,12671297	CONPC	5,74149402
RAWEDW	2,66161922	NGRCOOFR	4,64321777	DIFF	5,98850934
NGRUKKOR	2,78873311	ESFMODEK	4,68471819	CONSPCST	6,06599364
NGRSUMCF	2,80134352	RAWPCWST	4,95775163	CONSPC	6,2311675
HARMCORE	2,97785563	FOURRW	5,17372728	FOURRI	7,22128117
				0,5 KONST	8,81705615

Tabelle 4-24: Euklidische Distanzen der algorithmischen Ähnlichkeitsmaße zu den *ratings* der Vpn. über die 13 richtigen und falschen Varianten in Exp. 2.

Die besten Maße der fünf Dimensionen über alle 24 Melodievarianten des Kontrolltests waren: *diffed* (2,04), *ngrukkon* (2,44), *harmcore* (2,98), *consed* (3,57) und *rhythfuzz* (3,65). Die Distanzen reichten insgesamt von 2,04 bis 7,22 (die Kontrollkonstante 0,5 nicht berücksichtigt). Das beste Regressionmodell ergab sich mit den drei Maßen *ngrukkon*, *rhythfuzz* und *harmcore*.

Dieses Modell schätzte die Urteile der Musikexperten sehr gut: $R = 0,96$, $R^2 = 0,921$, korrigiertes $R^2 = 0,909$, Standardfehler der Schätzung 0,49.

Eine Kontrollrechnung mit den Maßen des optimalen Modells für die richtigen Varianten (*rawedw* und *ngrcoord*) erreichte deutlich schlechtere Werte der Anpassungsgüte: $R = 0,917$, $R^2 = 0,841$, korrigiertes $R^2 = 0,826$ und Standardfehler der Schätzung = 0,68.

Die optimale Regressionsgleichung zur Erklärung der Urteile der Versuchspersonen zu allen Melodievarianten des Kontrollexperiments ist:

$$\hat{y} = 3,027 \cdot ngrukkon + 2,502 \cdot rhythfuzz + 1,439 \cdot harmcore$$

In der Normierung auf den Bereich von 0 bis 1 ergibt sich die folgende Gleichung, die bei Identität der beiden zu vergleichenden Melodien einen Wert >1 liefert. Dies mag jedoch höchstens ein Problem der mathematische Ästhetik sein. Bei praktischen Anwendungen würde man einfachheitshalber allen Werten >1 den Ähnlichkeitswert 1 zuweisen.

$$\hat{y} = -0,146 + 0,505 \cdot ngrukkon + 0,417 \cdot rhythfuzz + 0,24 \cdot harmcore$$

Dieses kombinierte Maß (*opti3*) kam den Urteilen der Musikexperten sehr nahe: Die euklidische Distanz über alle 24 Melodievergleichspaare betrug 1,364; ein Wert, der um 33,4%

besser als der des besten Einzelmaßes *diffed* ist. Auch dieses Ergebnis lässt sich gut anhand der folgenden Abbildung darstellen.

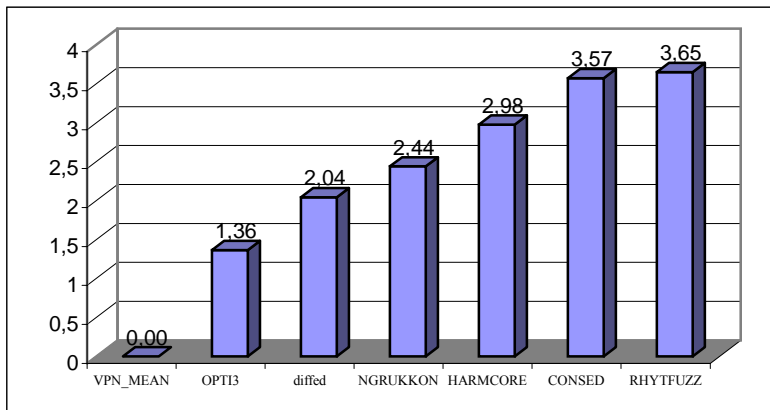


Abb. 4-15: euklidische Distanz der besten Ähnlichkeitsalgorithmen im Vergleich mit den mittleren *Ähnlichkeitsratings* der Vpn. über alle 24 Varianten von Exp.2.

Offensichtlich wird Information aus sehr unterschiedlichen Quellen (Unterschiede im Vorkommen kurzer Motive, Rhythmus und harmonisch gewichtete Töne) gebraucht, um die Urteile der Versuchspersonen zu modellieren.

Diese Ergebnis scheint sehr plausibel, wenn man unterstellt, dass Dimensionen wie Rhythmus, harmonischer Gehalt oder das Vorkommen einzelner Motive leichter und vielleicht sogar implizit ohne erlernte musikalische Fähigkeiten verarbeitet werden können. Die Auswertung dieser ‚einfachen‘ Dimensionen ist aber offensichtlich schon für die Bewältigung der Aufgabe des Kontrollexperiments ausreichend. Genauso plausibel ist, dass das Ukkonen-Maß das beste Ergebnis aller *n-gram*-Maße zeigte, da dieses die Unterschiede zwischen den Häufigkeiten der Substrings zweier Symbolsequenzen zählt.

Insgesamt betrachtet scheint es so, dass Musikexperten ihre Urteilsstrategie für Ähnlichkeitsbeurteilungen an den Kontext der jeweiligen Aufgabe und experimentellen Stimuli anpassen. Für die subtileren Unterschiede zwischen Varianten des selben Ursprungs werden eher die Tonhöhen- bzw. Intervallldimension genutzt, während bei der Ähnlichkeitsbewertung von falschen und richtigen Varianten eher harmonische und Rhythmusinformationen sowie das Vorhandensein von unterschiedlichen Motiven beachtet werden.

Die zweite der eingangs dieses Abschnittes gestellten Fragen ist vor allem für Algorithmen von Bedeutung, die Melodievarianten aufgrund von Ähnlichkeiten einer unter mehreren Referenzmelodien zuordnen müssen. Dies ist in der Regel bei *Melodie-Retrievalsystemen*, wie sie beispielsweise für *Query-by-humming*-Anwendungen konstruiert werden, der Fall. Diese Frage wurde auf zwei Arten operationalisiert:

Zum ersten wurde in einer Reihe von t-Tests diejenigen Maße identifiziert, bei denen die mittlere Ähnlichkeit der richtigen Varianten in beiden Melodien des Kontrolltests signifikant ($p < 0,05$) größer war als die mittlere Ähnlichkeit der falschen Varianten. Dies war für die Maße *Vpn_mean* (Mittel der Urteile der Musikexperten), *rawed*, *diffed*, *ngrukkon*, *rhythgaus* und *rhythfuzz* der Fall.

Zum zweiten wurde mittels eines U-Tests nach Mann und Whitney die Zahl der Inversionen, d.h. der Vertauschungen von falschen und richtigen Varianten bestimmt. Der zugehörige Signifikanztest gibt an, ob sich diese Vertauschungen bei einer gegebenen Irrtumswahrscheinlichkeit ($p < 0,05$) auch zufällig hätten ergeben können, wenn kein Unterschied zwischen den beiden Gruppen bestünde. Melodie A besaß 5 falsche und 7 richtige Varianten, Melodie C jeweils 6 Varianten. Die besten Maße gibt die folgende Tabelle wieder:

Maß	Anzahl Inversionen Melodie A	Anzahl Inversionen Melodie C	Summe Inversionen
Vpn_mean	0 ($p=0,0015$)	0 ($p=0,001$)	0
Opti3	0 ($p=0,0015$)	0 ($p=0,001$)	0
Rhythgaus	1 ($p=0,0025$)	3 ($p=0,0075$)	4
Diffed	2,5 ($p=0,005$)	3 ($p=0,0075$)	5,5
rhythfuzz	0 ($p=0,0015$)	6 ($p=0,0325$)	6
Ngrukkon	7 ($p=0,05$)	0 ($p=0,001$)	7
Rawed	3 ($p=0,009$)	6 ($p=0,0325$)	9

Tabelle 4-25: Anzahl der Vertauschungen der Reihenfolge der Ähnlichkeitswerte von richtigen und falschen Varianten durch die 5 besten Ähnlichkeitsmaße, die Versuchspersonen und das optimierte Maß *opti3*.

Wie leicht zu erkennen ist, lieferte keines der einzelnen algorithmischen Maße eine fehlerfreie Trennung von richtigen und falschen Varianten. Das kombinierte Maß, welches aus der Optimierung hervorging (*opti3*), zeigt hingegen die gleiche Trennschärfe zwischen richtigen und falschen Varianten wie die Urteile der Versuchspersonen.

Erstaunlich ist, dass die beiden Rhythmusmaße *rhythgaus* und *rhythfuzz* bei dieser Aufgabe sehr gut abschneiden, bedenkt man noch deren geringe Bedeutung bei der Schätzung der Ähnlichkeit im Hauptexperiment. Offensichtlich kann die Dimension Rhythmus gut dazu dienen, richtige Melodievarianten von falschen recht verlässlich zu unterscheiden.

Zu dem inhaltlich fast gleichen, numerisch aber unterschiedlichen Ergebnis kommt der Vergleich der Flächen unter den Receiver Operating Characteristic-Kurven (ROC). Diese Kurven und das zugehörige AUC-Maß wurden in Kapitel 3.2 ausführlich diskutiert.

Analog zu den gerade dargestellten Ergebnissen des U-Tests ergeben sich bei der Berechnung der Area-scores (AUC) und der zugehörigen Konfidenzintervalle auf dem 95%-Niveau die folgenden Werte für die optimalen Maße aus Experiment 2:

Maß	AUC (Melodie A)	Konfidenzintervall (Melodie A)	AUC (Melodie C)	Konfidenzintervall (Melodie C)
Vpn_mean	1	1 - 1	1	1 - 1
Opti2	1	1 - 1	1	1 - 1
Rhythgaus	0,96	0,78 - 1	0,9	0,6 - 1
Diffed	1	1 - 1	0,92	0,56 - 1
rhythfuzz	1	1 - 1	0,83	0,45 - 1
Ngrukkon	0,89	0,79 - 1	1	1 - 1
Rawed	0,91	0,64 - 1	0,86	0,56 - 1

Tabelle 4-26: AUC-Werte für die Diskriminierung von richtigen und falschen Varianten durch die 5 besten Ähnlichkeitsmaße, die Versuchspersonen und das optimierte Maß opti2.

Da die Berechnung der *area scores* auch die Information des Intervallskalenniveaus der Ähnlichkeitsmaße ausnutzt, d.h. die Distanzen zwischen den Ähnlichkeitswerten, ergibt sich ein leicht verschiedenes Ergebnis, betrachtet man die Performanz der einzelnen Maße anhand der Area scores. So würde nach dieser Tabelle beispielsweise *diffed* vor *rhythgaus* in einem Ranking der Ähnlichkeitsmaße liegen. Insgesamt liegen sowohl die fünf besten Maße als auch natürlich *opti2* in dem Bereich zwischen 0,8 und 1, der in der Literatur generell als sehr gut („excellent“ für AUC > 0,8 bzw. „outstanding“ für AUC > 0,9 bei Hosmer & Lemeshow (2000, S. 162)) beurteilt wird. Eine größere Melodiestichprobe als die getesteten 2x12 Varianten würden hier sicher ein differenzierteres Ergebnis hinsichtlich der Diskriminationsleistung der getesteten Maße liefern.

Die relevanten Plots aus dem frei verfügbaren Programm *ROC_AUC* (von Boris Schröder vom Institut für Geoökologie der Universität Potsdam) zum *area score* und der ROC-Kurve sind in zwei Grafiken im Anhang dargestellt.

4.4.3 Ermittlung des optimalen Ähnlichkeitsmaßes für das zweite Kontrollexperiment (Experiment 3)

Wie in den beiden vorangegangenen Abschnitten wurde auch für die Daten des zweiten Kontrollexperimentes ein optimales Ähnlichkeitsmaß mit Hilfe der linearen Regression unter Einschluß der besten Einzelmaße der fünf inhaltlichen Dimensionen gesucht.

Als beste Maße ergaben sich (euklidische Distanz zu *Expertenratings* in Klammern): *ngcoorf* (3,34), *rawed* (3,50), *coned* (3,67), *harmcore* (4,05) und *rhythgaus* (5,94).

Das beste Regressionsmodell nahm die Maße *ngcoorf* und *rawed* auf und erreichte Werte für die Modellgüte von $R^2 = 0,803$ und korrigiertem $R^2 = 0,79$, mit einem Standardfehler der Schätzung von 0,98.

Unter Verwendung der beiden Einzelmaße, die sich in Experiment 1 für das optimale Modell gefunden hatten – *rawedw* und *ngrcoord* – ergab sich ein leicht schlechteres Modell: $R^2 = 0,791$ und korrigiertem $R^2 = 0,776$, mit einem Standardfehler der Schätzung von 1,01.

Umgekehrt ergab sich ein deutlich schlechteres Modell mit den Regressoren *ngcoorf* und *rawed* für die Daten von Experiment 1 ($R^2 = 0,76$; korrigiertes $R^2 = 0,704$, Standardfehler der Schätzung: 0,787) verglichen mit dem oben beschriebenen optimalen Modell für diese Daten. Die euklidische Distanz zwischen *optil* und den Mittel der Versuchspersonenurteile betrug 2,633. Dieser Wert ist um 21,17% besser als der des besten Einzelmaßes (*ngcoorf*: 3,34). Auch in diesem Fall stellt das Regressionsmodell also eine echte Optimierung dar.

So erweist sich das in Experiment 1 gefundene optimale Modell aus *rawedw* und *ngrcoord* für die drei verschiedenen Datensätze aus den Experimenten 1-3, die sich für Ähnlichkeitsbeziehungen zwischen Varianten derselben Melodie ergeben hatten, insgesamt am leistungsfähigsten.

Da in Experiment 3 wie in Experiment 1 der Kontext für die Experten nur aus richtigen Varianten bestand, konnten die Gewichte für das auf den Bereich von 0 bis 1 normierte Maß beibehalten werden. Es ergab sich eine Korrelation zwischen *optil* und den gemittelten Urteilen der Experten von 0,888. Der hohe Zusammenhang lässt sich auch folgender Grafik entnehmen (*vpmea301* steht dabei für das gemittelte Urteil der Experten im 3. Experiment und auf den Bereich 0 bis 1 bezogen; *optil03* bezeichnet das Maß *optil* bezogen auf den Bereich 0 bis 1):

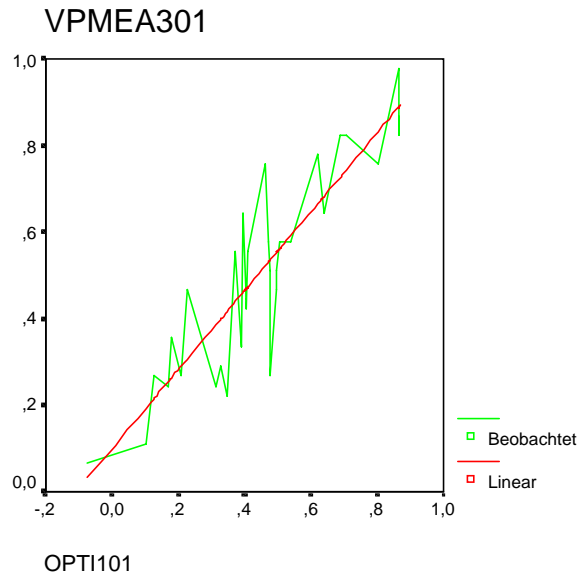


Abb. 4-16: Zusammenhang zwischen den mittleren *Ähnlichkeitsratings* der Vpn. und dem kombinierten Ähnlichkeitsmaß opti1 für die 32 Melodievarianten aus Exp. 3.

4.4.4 Allgemeine Beobachtungen zum Verhalten und der Leistungsfähigkeit der getesteten Ähnlichkeitsmaße

Abgesehen von der Ermittlung der besten Ähnlichkeitsmaße und des optimalen Maßes können die Distanzen der getesteten Maße zu den Ähnlichkeitsurteilen der Musikexperten auch Auskunft darüber geben, welche Datentransformationen und Ansätze zur Ähnlichkeitsmessung generell effektiv und sinnvoll sind. Auch die Frage, welche Datentransformationen zusammen mit welchen Ähnlichkeitsmaßen die besten Ergebnisse liefern, lässt sich anhand der Analyse der euklidischen Distanzen (s. Tabellen oben) zu den *Versuchspersonenratings* bzw. durch die Analyse der Rangplätze in diesen Tabellen beantworten. Die folgenden Beobachtungen gehen dabei von den euklidischen Distanzen des Hauptexperiments aus, sofern nichts anderes angegeben ist.

Kontur vs. rohe Tonhöhe:

Generell kamen Konturmaße den Urteilen der Versuchspersonen näher als die analogen Maße, die auf den rohen Tonhöhenwerten basierten. Dies trifft sowohl für die Maße der Edit Distance als auch auf die Korrelationsmaße zu, jedoch nicht auf die *n-grams*, die Intervallrichtungen (=Kontur) verwendeten. Die Distanzen zu den *ratings* der Versuchspersonen waren jeweils für die besten Konturversionen der Maße geringer:

coned < rawedw;
conpc < rawpc;
conpcst < rawpcstw

Vergleich der verschiedenen Ähnlichkeitsansätze:

Zur Evaluierung der verschiedenen Ähnlichkeitsmaße (auf Basis von Tonhöhen- und Intervalldaten) wurde ein nicht-parametrischer Kruskal-Wallis-Test auf den Rängen der Distanzen zu den *Versuchspersonenratings* gerechnet. Der Test ergab einen signifikanten Wert mit einer Irrtumswahrscheinlichkeit von $p=0,005$. Die Maße der Edit Distance zeigten die besten Resultate mit einem mittleren Rang von 3, während die *n-gram*-Maße einen mittleren Rang von 6 erreichten. Korrelations- und Differenzmaße blieben mit mittleren Rängen von je 13,5 deutlich dahinter zurück. Die Edit Distance ist also das effektivste Verfahren zum Vergleich von Melodievarianten, dicht gefolgt von den *n-gram*-Maßen. Einen ebenfalls nur leichten Vorteil des Edit Distance-Ansatzes vor den besten *n-gram*-Maßen fanden auch Uitdenbogerd und Zobel (2002), obwohl sie einen grundverschiedenen experimentellen Ansatz und vollkommen andere Teststimuli verwendeten.

Betrachtet man die verschiedenen Ähnlichkeitsmaße auf Basis der Konturdaten, so ergibt sich ein leicht anderes Bild. Zwar erreichten wieder die Edit Distance-Maße die besten Resultate (mittlerer Rang 1,5), während Korrelations- und Fouriermaße mit einem mittleren Rang von je 8 in der Mitte lagen. Die *n-gram*-Maße, die nur die Intervallrichtungen verarbeiteten (Parsons Code), zeigten die schlechtesten Ergebnisse (mittlerer Rang 12,33). Der Kruskal-Wallis-Test kam insgesamt mit einem p -Wert von 0,071 der üblichen Signifikanzgrenze sehr nahe, ohne freilich signifikant zu werden.

Definitionen melodischer Kontur:

Was die beiden Definitionen von melodischer Kontur angeht, die in dieser Studie verwendet wurden, so war die Definition von Müllensiefen und Frieler der Steinbeckschen bei allen drei getesteten Maßen überlegen, was sich an den strikt geringeren Distanzen zu den *ratings* der Musikexperten festmachen läßt:

coned<consed
conpest<conspcst
conpc<conspc

n-gram-Ansätze:

Unter den drei verschiedenen *n-gram*-Ansätzen waren über alle vier Typen von Ausgangsdaten hinweg *Coordinate Matching* und das Ukkonen-Maß gleich erfolgreich (mittlerer Rang je 4,5), während das *Sum-Common*-Maß in jedem Fall als schlechtestes abschnitt (mittlerer Rang 10,5). Der Kruskal-Wallis-Test war mit einer Irrtumswahrscheinlichkeit von 0,016 signifikant. Auch dieses Ergebnis stimmt in der Tendenz mit der vorgenannten Studie von Uitdenbogerd und Zobel (2002) überein, die *Coordinate Matching* als bestes Globalmaß unter den *n-gram*-Ansätzen ausmachten.

Im Hauptexperiment waren die *n-gram*-Maße, die auf Intervalldaten beruhten weit effektiver als die Maße mit klassifizierten Intervallen und Intervallrichtungen als Ausgangsmaterial. Die generelle Überlegenheit der *n-gram*-Maße auf Intervalldaten zeigte sich auch in den Daten des Kontrollexperiments.

In allen getesteten Fällen gaben *n-grams* der Länge 3 die besten Ergebnisse. Da der hier implementierte Algorithmus immer den maximalen Ähnlichkeitswert aller *n-gram*-Längen als ‚wahren‘ Ähnlichkeitswert wählte, ist dieses Ergebnis nicht überraschend, denn bei Zeichenfolgen aus einem begrenzten Alphabet ist die Wahrscheinlichkeit für kürzere Zeichenketten für eine Übereinstimmung größer. Dass aufgrund der *3-grams* die Ähnlichkeit dennoch nicht überschätzt wurde, zeigt die generell sehr gute Performance aller *n-gram*-Maße zumindest auf Intervallbasis. Diese Überlegenheit der kurzen *n-grams* mag auch mit der Länge der hier verwendeten Testmelodien von 30 bis 50 Tönen zusammenhängen. Bei längeren Testmelodien dürften auch längere *n-grams* überlegen sein, wie Uitdenbogerd und Zobel (2002) belegen.

Rhythmusmaße:

Unter den eingesetzten Rhythmusmaßen zeigte *rhytfuzz* die größte Nähe zu den *ratings* der Musikexperten in den Daten beider Experimente. Das gaussifizierte Rhythmusmaß nach Frieler wiederum ergab bessere Resultate als alle drei *n-gram*-Rhythmusmaße. Von diesen war das Ukkonen-Maß für die Rhythmusdaten das erfolgreichste.

Begrenzung des Korrelationskoeffizienten auf den Wertebereich 0 bis 1:

Bei allen Korrelationsmaßen mit Ausnahme der Fouriermaße waren die Varianten, bei denen der Koeffizient auf den Wertebereich von 0 bis 1 beschränkt war, denen überlegen, die den üblichen Wertebereich (-1 bis 1) ausnutzen:

$$\text{conpcst} > \text{conpc}$$

$$\text{conspcst} > \text{conspc}$$

$$\text{rawpcst} > \text{rawpc}$$

$$\text{rawpcwst} > \text{rawpcw}$$

Dies lässt sich plausibel damit begründen, dass für die Versuchspersonen eine entgegengesetzte oder inverse Ähnlichkeit bei der Beurteilung der Varianten keine Rolle spielte.

Gewichtung an Tondauern:

Eine letzte Beobachtung betrifft die Gewichtung von Tonhöhenmaßen anhand der Tondauern, wie sie beispielsweise von Steinbeck vorgeschlagen und oben ausführlich erläutert wurde. Im Vergleich aller Ähnlichkeitsmaße, die eine gewichtete und ungewichtete Variante besitzen, ergab sich kein Vorteil für eine der beiden Optionen. Sie erreichten beide den selben Rang in einem U-Test. Es lässt sich also kein generelles Urteil abgeben, ob allgemein die Gewichtung an den Tondauern für Tonhöhenmaße verwendet werden sollte. Auf der Ebene der einzelnen Maß-Familien jedoch lässt sich eine solche Unterscheidung treffen: Die Edit Distance der rohen Tonhöhen (*rawed* bzw. *rawedw*) schnitt besser mit der Dauergewichtung ab, während die Korrelationsmaße für rohe Tonhöhen in der Tendenz ohne die Gewichtung bessere Resultate erbrachten. (Alle Konturmaße wurden mit Dauergewichtung implementiert, sodass für diese Gruppe von Maßen kein Vergleich der beiden Optionen möglich war) Die Fouriermaße schließlich schnitten leicht besser in den gewichteten Varianten ab.

4.5 Vergleich der Ähnlichkeitsmaße in Multidimensionaler Skalierung

4.5.1 Ansatz und Methode

Ein abschließender Vergleich der verschiedenen hier implementierten Ähnlichkeitmaße soll eine Gruppierung und Einordnung in verschiedene Dimensionen liefern. Die Zuordnung der verwendeten Ähnlichkeitsmaße zu den fünf Dimensionen, die weiter oben erläu-

tert wurde, war eine heuristische ‚ad-hoc-Methode‘, die von den verwendeten Ausgangsdaten, einer Betrachtung der Algorithmen und den Intentionen der Forscher ausging, die die jeweiligen Maße in früheren Studien eingesetzt hatten.

Durch diese ad-hoc-Methode ist jedoch keinesfalls gewährleistet, dass die verschiedenen Ähnlichkeitsmaße wirklich Unterschiede bzw. Information auf den behaupteten Dimensionen messen. Es drängen sich eine Reihe von Fragen hinsichtlich dieser Systematik auf:

Ist es wirklich so, dass die Fouriermaße Periodizitäten in den Tonhöhenfolgen von Melodien und damit deren Konturen vergleichen, oder ist dies nur eine Metapher, die Schmuckler benutzt, um diesem Maß eine Legitimation in seiner Studie zu geben?

Lässt sich wirklich behaupten, die n-gram-Maße auf Basis der Intervallrichtungen würden die Kontur messen, wo Kontur doch vielleicht eher durch längere Tendenzen von Auf- und Abbewegungen wahrgenommen und charakterisiert werden sollte?

Gerade die Tatsache, dass in den optimalen Regressionsmodellen an keiner Stelle ein Tonhöhen-/Intervallmaß und ein Konturmaß gleichzeitig eingehen, lässt stark vermuten, dass die Maße von beiden Dimensionen eigentlich eine gemeinsame Größe messen.

Mißt *harmcore* wirklich die impliziten Harmonien einer Melodie oder ist es doch besser als Tonhöhenmaß zu klassifizieren?

Steht die Dimension „Motive / kurze Elementfolgen“ wirklich orthogonal auf den übrigen Dimensionen oder besteht ein Zusammenhang mit der Tonhöhendimension?

Eine elegante Methode zur Beantwortung dieser Fragen stellt die Multidimensionale Skalierung (MDS) dar, die dazu geeignet ist, latente Dimensionen zu entdecken und zu skalieren, nach denen sich eine Menge von Objekten beschreiben lässt, zwischen denen sich Distanzen messen lassen.

Die Objekte stellen hier die verschiedenen Ähnlichkeitsmaße dar, zwischen denen sich über die 84 Vergleichpaare euklidische Distanzen nach der oben beschriebenen Formel berechnen lassen.

Die Multidimensionale Skalierung wurde in fünf Schritten durchgeführt:

a) Auswahl der Variablen

Um die MDS übersichtlich zu halten, wurden von den Ähnlichkeitsmaßen, die in unterschiedlichen Varianten vorlagen, jeweils die ausgewählt, die eine geringere Distanz zu den

Urteilen der Versuchspersonen besaßen, d.h. die bessere Resultate geliefert hatten. Dieser Selektionsprozeß betraf die Konturmaße (hier wurde jeweils die Müllensiefen-Frieler-Variante gewählt), die Korrelationsmaße (hier wurden nur die Varianten mit dem Wertebereich 0-1 gewählt), die *n-gram*-Maße für Intervallfolgen (es wurden die Maße mit exakten Intervallen gewählt) und die Gewichtung (hier wurde die jeweils bessere Variante einer Familie gewählt). Die für die MDS ausgewählten 19 Variablen finden sich in folgender Tabelle:

CONED	NGRUKKFR
CONPCST	NGRUKKON
DIFFED	OPT1
FOURR	RAWED
FOURRWST	RAWEDW
HARMCORE	RAWPCST
NGRCOOFR	RHYTFUZZ
NGRCOORD	RHYTGAUS
NGRSUMCO	VPN_MEAN
NGRSUMFR	

Tabelle: 4-27: Für die MDS verwendete Ähnlichkeitsmaße

b) Berechnung der Distanzen und z-Transformation

Alle Variablen wurden z-transformiert, um eine Vergleichbarkeit zwischen *Versuchspersonenratings* (theoretischer Wertebereich 1-7) und den Ähnlichkeitsmaßen (theoretischer Wertebereich 0-1) herzustellen.

Danach wurden die euklidischen Distanzen zwischen allen z-transformierten Variablen jeweils paarweise nach der schon oben im Abschnitt zur Optimierung der Ähnlichkeitsmaße beschriebenen Formel berechnet. Es ergab sich eine 19x19-Distanzmatrix. Diese diente als Ausgangsmatrix für die MDS.

c) Parameter für Ermittlung der Konfigurationen

Zur Berechnung der MDS-Konfiguration wurde der in SPSS implementierte ALSCAL-Algorithmus verwendet. Für die Transformation der Distanzen zwischen den Variablen in Disparitäten wurde ein ordinaler Zusammenhang vorausgesetzt. Trotz dieses ‚nur‘ ordinalen Zusammenhangs, der für die Ausgangsdaten und die Transformationen in Disparitäten vorgegeben wird, ist eine metrische MDS als Lösung möglich. Dies ist durch die hohe Datenverdichtung gewährleistet, die selbst noch für eine mögliche vierdimensionale Lö-

sung gegeben wäre. Als Datenverdichtungskoeffizient für eine vierdimensionale Lösung für die Anordnung der 19 Variablen ergibt sich:

$$Q = \frac{171 \text{Ähnlichkeitswerte}}{76 \text{Koordinaten}} = 2,25$$

Backhaus bezeichnet Werte mit $Q \geq 2$ als akzeptabel (vgl. Backhaus et al., 2003, S. 633).

d) Festlegung des Dimensionsraumes

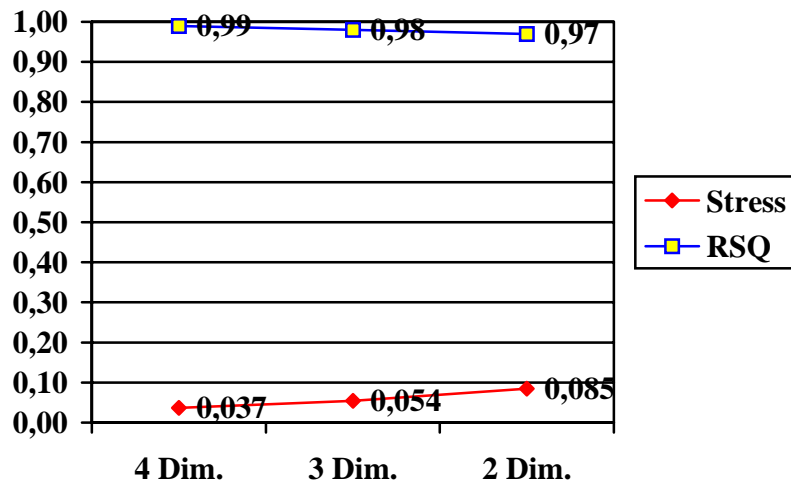
Zur Festlegung der Dimensionalität des Raumes, in dem die Konfiguration der 19 Ähnlichkeitsmaße beschrieben werden soll, sind zwei Modellgrößen ausschlaggebend, die die Anpassungsgüte des MDS-Modells im jeweiligen n-dimensionalen Raum repräsentieren sowie die inhaltliche Plausibilität der Interpretation der jeweiligen Lösung. Die beiden Größen sind der sogenannte Stress-Wert und der RSQ-Wert, die sich als Anpassung der Disparitäten an die Distanzen in der n-dimensionalen Konfiguration bzw. als quadrierte Korrelation zwischen Distanzen und Disparitäten interpretieren lassen. Der Stress-Wert sollte dabei möglichst gering und der RSQ-Wert möglichst nahe 1 sein. Für den Stress gelten gemeinhin Werte geringer als 0,1 als hervorragend und Werte zwischen 0,1 und 0,2 als akzeptabel. Für die inhaltliche Interpretation einer Lösung wurde vor allem darauf geachtet, dass die Ähnlichkeitsmaße, die a) jeweils ganz verschiedene Ausgangsdaten verwenden oder b) durch die Konstruktion des Algorithmus sehr wahrscheinlich verschiedene Aspekte zweier Melodien vergleichen, hohe Werte auf verschiedenen Dimensionen aufweisen. Der Fall a) wäre z.B. für *rhythfuzz* und *rawed* gegeben während Fall b) beispielsweise für *harmcore* und die Fouriermaße zutreffen dürfte.

e) Interpretation

Die Interpretation einer numerisch akzeptablen und auf den ersten Blick inhaltlich plausiblen Lösung wird zum einen durch die Darstellung der Konfiguration der Ähnlichkeitsmaße in Graphen unterstützt. Zum anderen kann ein Vergleich der Dimensionen mit den Fehlerzahlen der fünf verschiedenen Fehlertypen bei der Interpretation helfen. Dies wird über eine lineare Regression realisiert, die die Fehlerzahlen eines Typs durch die Werte als zu erklärende Variable definiert und die Koordinatenwerte einer Dimension als Regressor.

4.5.2 Durchführung der MDS

Mit den gerade genannten Parametern wurden mit SPSS vier MDS-Lösungen mit 4 bis 2 Dimensionen gerechnet. Einen Überblick, über die Entwicklung des Stress-Maßes und von



R^2 gibt die folgende Grafik:

Abb. 4-17: Entwicklung der Stress- und RSQ-Werte über die MDS-Lösungen unterschiedlicher Dimensionalität.

Wie deutlich zu sehen ist, nimmt der Stress-Wert in etwa linear zu und der RSQ-Werte linear ab. Vor der 2-dimensionalen Lösung ist kein ‚Knick‘ in der Zunahme des Stressses zu erkennen. Dies spricht für die 2-dimensionale Lösung. Mit einem Stress von 8,5% und einem RSQ von 0,97 ist diese Lösung nach den üblichen Kriterien als sehr gut oder hervorragend zu bezeichnen (vgl. Kruskal & Wish, 1978; Borg & Staufenbiel, 1993; Backhaus et al., 2003).

Die Koordinaten der Ähnlichkeitsmaße auf den drei Dimensionen gibt die folgende Tabelle wieder:

Ähnlichkeitsmaß	Dimension1	Dimension 2
RAWED	,5012	-,5130
RAWEDW	,3455	-,5497
RAWPCST	-,2264	-1,2317
CONED	,9222	-,0323
CONPCST	,3735	-1,0342
FOURR	-,6774	-,9789
FOURRWST	-,5950	-1,2030
DIFFED	,6983	-,3126
NGRSUMCO	1,3112	1,0678
NGRUKKON	1,4015	1,0170
NGRCOORD	1,4132	,9964
NGRSUMFR	-1,6504	1,3993
NGRUKKFR	-1,6778	1,1899
NGRCOOFR	-1,7772	,9668
RHYTGAUS	-1,6304	-,4102
RHYTFUZZ	-1,3936	-,0633
HARMCORE	,5954	-,6437
VPN_MEAN	1,1530	,2599
OPTI1	,9132	,0756

Tabelle 4-28: Koordinatenwerte der Ähnlichkeitsmaße in der 2-dimensionalen MDS-Lösung.

Auch die oben genannten oberflächlichen inhaltlichen Kriterien scheinen durch die Konfiguration erfüllt. So zeigen *rhythfuzz* und *rawed* sowie *harmcore* und die beiden Fouriermaße stark verschiedene Werte auf der 1. Dimension.

Auch das sogenannte Shepard-Diagramm zeigt die Güte der Lösung anhand der gegeneinander abgetragenen Disparitäten und Distanzen der MDS-Konfiguration, die sich gut um eine gedachte Gerade durch den Ursprung gruppieren.

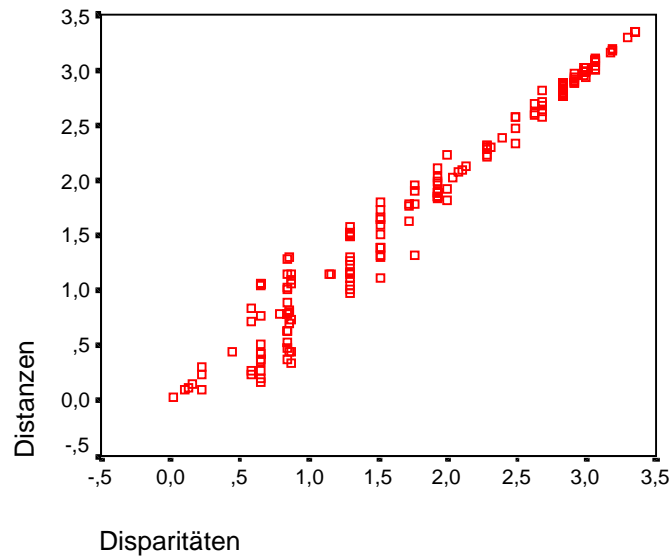


Abb. 4-18: Sheppard-Diagramm der ursprünglichen Distanzen der Ähnlichkeitsmaße und ihrer Disparitäten in der zweidimensionalen Lösung (Daten aus Exp. 1).

Die nächste Grafik zeigt die Positionen der Ähnlichkeitsmaße im zweidimensionalen Koordinatensystem. Zu beachten ist, dass die Bezeichnungen einiger Datenpunkte zum Zwecke der besseren Lesbarkeit der Grafik insgesamt weggelassen wurden. Dies betrifft die Maße *ngrukkon* und *ngrsumco*, die fast genau unter *ngrcoord* liegen sowie *rawed*, das sehr nahe an *rawedw* liegt und schließlich *fourrwst*, welches nahe bei *fourr* liegt.

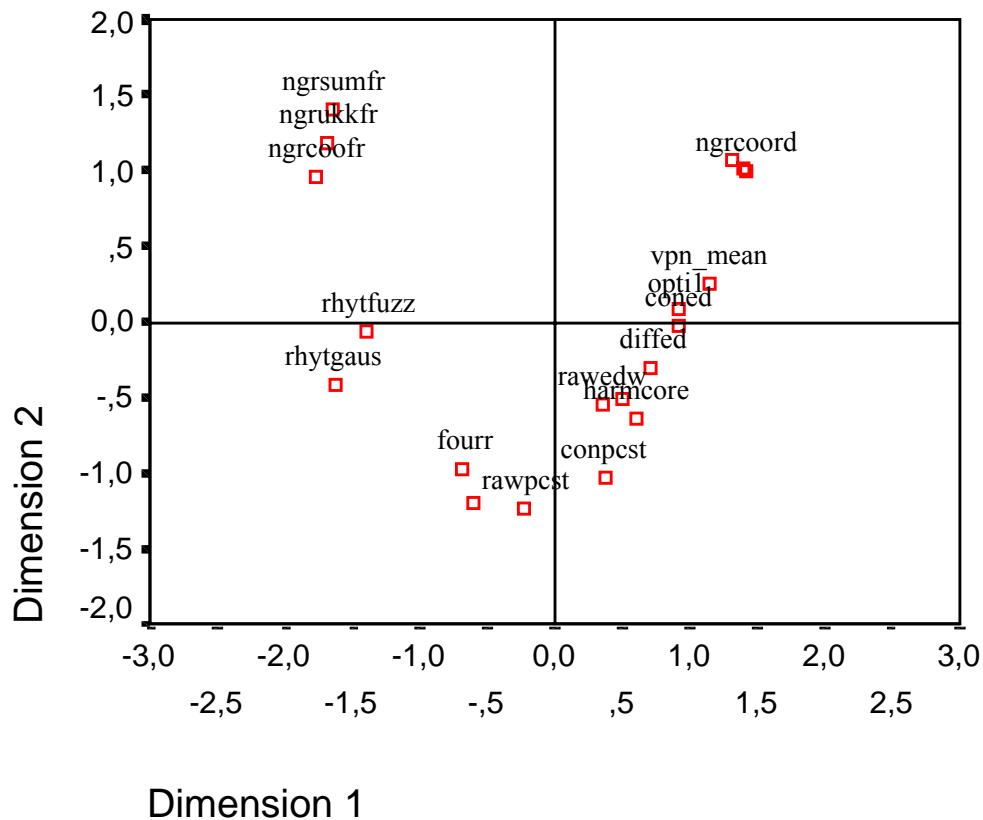


Abb. 4-19: Konfiguration der Ähnlichkeitsmaße im zweidimensionalen Raum (Daten aus Exp. 1).

Für die beiden Dimensionen der Konfiguration drängt sich anhand der Tabelle der Koordinaten und ihrer grafischen Umsetzung eine Interpretation auf:

Dimension 1 kann sehr plausibel für die Verwendung rhythmischer Information stehen. Der Pol ‚Rhythmus‘ würde dann am negativen Ende der Skala liegen, was durch sämtliche Rhythmusmaße, die stark negativ auf Dimension 1 verortet sind, nahegelegt wird. Am positiven Ende von Dimension 1 finden sich Maße, wie die n-gram-Maße für Tonhöhendaten oder *diffed*, die keinerlei Rhythmusinformation berücksichtigen. Wollte man Dimension 1 als zweipolige Dimension bezeichnen, so käme der Ausdruck ‚Rhythmus- vs. Tonhöheninformationen‘ in Frage.

Dimension 2 wird vor allem charakterisiert durch die n-gram-Maße, die die größten positiven Werte besitzen und die Korrelationsmaße *fourr*, *rawpcst* und *conpcst* im negativen Bereich der Skala. Hier scheint sich auszudrücken, wie die Maße die Menge der Elemente einer Melodie verarbeiten. Die n-gram-Maße – unabhängig davon, ob sie rhythmische oder Tonhöheninformationen benutzen – lösen die Melodie in kurze Sequenzen auf, deren Reihenfolge keine Rolle für den Ähnlichkeitswert spielt. Sie betrachten also nicht die Melodie als Ganzes, sondern als Summe ihrer Teile. Die Korrelations- und auch die Fouriermaße sind dagegen sensibel für den Gesamtverlauf einer Melodie. Dies findet sich in der Struk-

tur der Korrelationsmaße dadurch wieder, dass sie die Abweichungen der einzelnen Tonhöhen von der mittleren Tonhöhe der gesamten Melodien vergleichen. Auch die Fouriermaße betrachten die Gesamtheit der Konturverläufe, die durch die Amplitudenkoeffizienten ausgedrückt wird. Eine Änderung der mittleren Tonhöhe im Verhältnis zu den Einzeltönen, z.B. durch sehr große Sprünge, erbringt hier große Unterschiede. Die n-gram-Maße dagegen sind von einzelnen abweichenden Sprüngen kaum beeinflusst. Dimension 2 kann deshalb mit ‚globale vs. lokale‘ Ähnlichkeit belegt werden. Die obige ad-hoc-Charakterisierung der n-gram-Maße als Maße für Motive erfährt hierdurch eine Rechtfertigung.

Plausibel scheint auch die Position der Versuchspersonen zu sein, die wie schon oben gezeigt für ihre Ähnlichkeitsurteile im Kontext von Varianten nur einer Referenzmelodie nur wenig auf die rhythmischen Informationen zurückgreifen. Auf der Dimension 2 sind die Urteile der Versuchspersonen etwa in der Mitte nahe des 0-Punkts angesiedelt. Auch dies scheint plausibel, wenn man annimmt, dass Musikexperten sowohl den Globalverlauf einer Melodie bewerten als auch lokale Übereinstimmungen und Abweichungen wahrnehmen. Es verwundert deshalb auch nicht, dass das optimierte Ähnlichkeitsmaß *optil*, welches insgesamt den geringsten Abstand zu den Versuchspersonen zeigt, aus einer Kombination aus lokalem Maß (*ngrcoord*) und einem globaler orientierten Maß (*rawedw*) besteht, das die Reihenfolge der Tonhöhenelemente über die Gesamtheit einer Melodie berücksichtigt.

Die Durchführung der MDS ist gleichermaßen plausibel wie erkenntnisgewinnend gewesen. Denn statt der ad-hoc vorausgesetzten fünf oder drei Dimensionen (Tonhöhe, Rhythmus, Harmonik) ließen sich die Maße offensichtlich besser in nur zwei Dimensionen darstellen. ‚Rhythmus‘ und ‚globale/lokale Ähnlichkeit‘. Es zeigte sich also, dass die Dimensionen Kontur und Tonhöhe – wenn überhaupt – nur relativ geringe Unterschiede aufweisen und sich der harmonische Zusammenhang gut zu anderen durch Dauern gewichteten Tonhöhenmaße gesellt. Dies mag bei Verwendung von mehreren unterschiedlichen Maßen zum harmonischen Gehalt jedoch anders ausfallen.

Mit diesem Wissen um die wichtigen Unterscheidungen, die die zwei gefundenen Dimensionen treffen, lässt sich bei der Konstruktion und Optimierung von zukünftigen Ähnlichkeitsmaßen gezielter vorgehen.

4.6 Zusammenfassung

Dieses vierte Kapitel stellt innerhalb der vorliegenden Arbeit eine quasi eigenständige Studie dar, in dem die menschliche Beurteilung melodischer Ähnlichkeit und ihre Modellierung durch Algorithmen erforscht werden. Diese Untersuchung der menschlichen Ähnlichkeitswahrnehmung wurde notwendig durch das Fehlen einer umfassenden vergleichenden Studie zu Verfahren der Messung melodischer Ähnlichkeit und deren kognitiver Angemessenheit in der Literatur. Wie zu Beginn des Kapitels dargelegt wurde, ist ein kognitiv adäquater Formalismus zur Bestimmung der Ähnlichkeit von Melodien ein sehr wertvolles Hilfsmittel bei der Analyse von Gedächtnisleistungen für Melodien, wie beispielsweise im folgenden Kapitel zu erkennen sein wird. Die Erkenntnisse dieses Kapitels lassen sich in vier Bereiche gliedern.

Homogenität der Ähnlichkeitsbeurteilungen von Experten:

Sowohl die Daten des Hauptexperiments wie auch die der beiden Kontrollexperimente wurden nach demselben Verfahren ausgewertet. Zwei Kriterien mussten die Urteile einer Vpn. hierzu erfüllen, um in der weiteren Datenanalyse verarbeitet zu werden. Die Beurteilung desselben Melodie-Variantenpaares durfte sich über die Zeit nicht bedeutsam verändern und Varianten, die zur Referenzmelodie identisch waren, mussten mindestens als sehr ähnlich beurteilt werden. Weniger als die Hälfte der Versuchspersonen konnten diese individuelle Urteilkonsistenz erlangen. Bei den ausgewählten Versuchspersonen zeigte sich in allen drei Experimenten, dass sie untereinander extrem ähnlich urteilten. Dieses homogene Ähnlichkeitsurteil der Experten wurde mit sehr hohen Werten diverser Zusammenhangsmaße belegt. Offensichtlich hatten die selektierten Versuchspersonen eine sehr ähnliche oder vielleicht sogar identische Vorstellung davon, was Ähnlichkeit zwischen kurzen Melodien bedeutet. Ohne grossen Informationsverlust konnten deswegen die Urteile der selektierten Versuchspersonen gemittelt werden. Da es sich bei den ausgewählten Versuchspersonen um Studierende der Musikwissenschaft handelt, die durchweg einen sehr aktiven musikalischen Hintergrund aufweisen, wurden die gemittelten Urteilswerte als ‚Expertenurteile‘ bezeichnet.

Das sehr homogene Urteilsverhalten der Musikexperten ist die Voraussetzung für die Modellierung der Urteile anhand der bekannten experimentellen Parameter und mit Hilfe der verschiedenen Ähnlichkeitsalgorithmen.

Modellierung der empirisch erhobenen Ähnlichkeitsurteile anhand der experimentellen Parameter:

Innerhalb mehrerer linearer Regressionsanalysen wurde versucht, die Ähnlichkeitsurteile der ausgewählten Versuchspersonen vorherzusagen. Die experimentellen Parameter waren die Anzahl der Abweichungen zwischen Referenzmelodie und Variante in jeder der fünf manipulierten Dimensionen (Rhythmus, Intervalle, Kontur, Reihenfolge und Modulation), die Position der Abweichungen in jeder Dimension und die Fehlerdichte in jeder Dimension.

Es konnte ein Modell gefunden werden, welches 82% der Varianz in den Urteilen der Experten erklärt und dessen Vorhersagen sehr hoch mit den Expertenurteilen korrelieren ($r = 0,907$). Dieses Modell verwendet neben den Fehlerzahlen in den fünf Dimensionen auch die an der Fehlerdichte gewichtete mittlere Fehlerposition der Intervallfehler. Die mittlere Position und Fehlerdichte in den übrigen Dimensionen spielt – zumindest bei den hier getesteten relativ kurzen Melodien – keine Rolle. Den größten Einfluss haben Fehler, bei denen die Reihenfolge der Phrasen vertauscht wird, solche, die einen Teil der Melodie auf eine andere Tonstufe verrutschen lassen (hier terminologisch nicht ganz sauber mit Modulation bezeichnet) sowie Intervall- und Konturfehler. Ein deutlich geringeres Gewicht besitzen die Rhythmusfehler und die an der Dichte gewichtete mittlere Fehlerposition der Intervallfehler.

Auf Grund der guten Werte dieses Vorhersagemodells lässt sich interpretieren, dass die Experten für ihr Ähnlichkeitsurteil tatsächlich die Anzahl der Abweichungen zwischen Original und Variante (unbewusst) zählen und gewisse Fehlertypen stärker gewichten als andere. Das lineare Modell kann als plausible wie hinreichend genaue Erklärung der kognitiven Ähnlichkeitsurteile – zumindest unter diesen experimentellen Bedingungen – gelten. Das Erklärungsmodell besitzt weit bessere Modellwerte als ein vergleichbares Modell, welches McAdams und Matzkin (2001) zur Erklärung ihrer experimentellen Befunde formuliert haben.

Modellierung der empirisch erhobenen Ähnlichkeitsurteile mit Hilfe der Algorithmen zur Ähnlichkeitsmessung:

Noch interessanter als die Modellierung der Ähnlichkeitsurteile mit den experimentellen Parametern ist im Zusammenhang dieser Arbeit eine Modellierung unter Verwendung der

im ersten Teil des Kapitels ausführlich beschriebenen Ähnlichkeitsalgorithmen. Denn erst durch ein Modell, welches auf robusten Algorithmen beruht, die auch in Unkenntnis der tatsächlichen Fehler oder Abweichungen bei einem gegebenen Melodie-Variantenpaar operieren können, wird es möglich, die melodische Ähnlichkeit in einer Vielzahl von experimentellen oder Anwendungssituationen als Analyseinstrument einzusetzen.

Die drei durchgeführten Experimente repräsentieren unterschiedliche Aufgaben, in denen ein Ähnlichkeitsurteil von den Experten gefragt war. Es ergaben sich konsequenterweise drei unterschiedliche optimale Ähnlichkeitsmodelle, die wiederum mit der Technik der linearen Regression ermittelt wurden. Musikexperten scheinen also je nach Aufgabe die Strategie bei ihren Ähnlichkeitsurteilen anzupassen und teilweise unterschiedliche melodische Dimensionen zu beachten oder verschieden zu gewichten. Beim Einsatz eines Ähnlichkeitsmodells in einer speziellen Anwendungssituation ist also zunächst der Bewertungskontext zu analysieren und danach das geeignete Ähnlichkeitsmodell zu wählen.

Das anhand der Daten von Experiment 1 ermittelte Ähnlichkeitsmodell erwies sich auch auf einem vergleichbaren Teildatensatz von Experiment 2 und auf den Daten von Experiment 3 als gutes Vorhersagemodell und kann somit als robust gelten. Auf den Daten von Experiment 1 erklärt es 83% der Varianz, was als hinreichend gut betrachtet werden kann.

Anhand der Daten von Experiment 2 wurde ein optimales Ähnlichkeitsmodell ermittelt, das für das in dieser Arbeit zentrale *Query-by-humming*-Szenario geeignet erscheint, da sich im Beurteilungskontext der Versuchspersonen Melodievarianten sowohl der Originalmelodie („richtige“ Varianten) als auch solche von anderen Melodien („falsche“ Varianten) fanden. Dieses optimale Ähnlichkeitsmodell berücksichtigte neben den Dimensionen Rhythmus und impliziter Harmonik auch Unterschiede zwischen den kurzen Motiven zweier Melodien, die durch die Intervalle gebildet werden. Dieses Modell zeigte eine ebenso gute Unterscheidungsleistung zwischen richtigen und falschen Melodievarianten wie die Versuchspersonen selbst. Kein anderes der getesteten algorithmischen Maße erreichte dieselbe Klassifizierungsqualität. In diesem Sinne hat sich die Optimierung der Ähnlichkeitsmaße durch Kombination verschiedener Maße als sehr erfolgreich erwiesen.

Beim Vergleich der 24 getesteten Ähnlichkeitsmaße und der dahinterstehenden Ansätze zur Ähnlichkeitsmessung ergaben sich eine ganze Reihe von Ergebnissen, die für im Forschungsbereich des sogenannten *Music Information Retrieval* auf Interesse stoßen dürften. Die Ergebnisse sind detailliert im Abschnitt 4.4.4 aufgelistet. Hervorzuheben wäre hier die Edit Distance, die sich über alle Situationen hinweg als effizientestes Vergleichsverfahren

erwies. Desweiteren ist die Überlegenheit der Maße der melodischen Kontur gegenüber den analogen Intervallmaßen hervorzuheben, die sich deutlich zeigte.

Multidimensionale Skalierung des Raums der Ähnlichkeitsmaße:

Da die ursprüngliche Einteilung der 24 verwendeten Ähnlichkeitsalgorithmen in inhaltliche Gruppen zunächst durch die entsprechenden Zuschreibungen zu den Maßen in der Literatur vorgenommen wurde, sollte schliesslich noch geklärt werden, ob diese Einteilung gerechtfertigt war oder ob die verschiedenen Algorithmen vielleicht andere inhaltliche Dimensionen messen als ihre Schöpfer annehmen.

Aus den Daten von Experiment 1 wurde eine Distanzmatrix berechnet und die unterschiedlichen Ähnlichkeitsmaße sowie die Versuchspersonen in einem multidimensionalen Raum skaliert. Eine zweidimensionale Lösung erwies sich als hinreichend gut. Die beiden Dimensionen der Lösung ließen sich als ‚Verwendung rhythmischer Information‘ und ‚globale vs. lokale Ähnlichkeit‘ (d.h. gesamter Melodieverlauf vs. Übereinstimmung in kurzen Motiven) interpretieren.

Die Versuchspersonen berücksichtigen nach ihrer Position in diesem Lösungsraum globale und lokale Ähnlichkeit ungefähr gleich stark, während sie – zumindest in Experiment 1 – weniger stark auf rhythmische Information in den Melodien zurückgreifen. Diese Positionierung der Versuchspersonen deckt sich mit der inhaltlichen Interpretation des optimalen Ähnlichkeitsmodells für Experiment 1.

5 Wiedererkennen unbekannter Melodien

5.1 Vorüberlegungen und Modellierungsrahmen

Das optimierte Maß für die Bestimmung der melodischen Ähnlichkeit, welches im vorangegangenen Kapitel ermittelt wurde, ist primär ein Hilfsmittel zur Beschreibung der Gedächtnisrepräsentation, die sich eine Person von der Melodie eines Musikstückes bildet. Wäre die im Gedächtnis repräsentierte Melodie direkt beobachtbar, so ließe sich wahrscheinlich ein Zusammenhang zwischen verschiedenen unabhängigen Variablen und der melodischen Ähnlichkeit als abhängige Variable beobachten. Zu den unabhängigen Variablen, die vermutlich einen Einfluss auf die Ähnlichkeit zwischen Gedächtnisrepräsentation und einer präsentierten Melodie haben, gehören u.a. die Anzahl der Wiederholungen der Melodie als Versuchstimulus (z.B. Zielinska & Miklaszewski, 1992) bzw. beim Hören in echten Lebenssituationen (z.B. Levitin, 1994), die Dauer des Retentionsintervalls (z.B. Dewitt & Crowder, 1986), für die die Melodie im Gedächtnis behalten werden muss und die strukturelle Einfachheit bzw. Komplexität der zu behaltenden Melodie (z.B. Boltz & Jones, 1986). Anhand der melodischen Ähnlichkeit lässt sich feststellen, wie genau oder wie gut die Repräsentation einer Melodie im Gedächtnis ist. Nun liegt die crux der Gedächtnisforschung in der Tatsache, dass sich Gedächtnisinhalte nicht direkt beobachten lassen, sondern nur Schlüsse auf ein hypothetisches Gedächtnissystem aus Daten aus Verhaltensexperimenten gefolgert werden können. Für den Bereich von Melodien bieten sich zur Ermittlung aussagekräftiger Daten grundsätzlich die beiden experimentellen Paradigmen von Wiedererkennen (*recognition*) und Wiedergabe (*recall*) an, die bereits in den Kapiteln 3.2 und 3.3 ausführlich diskutiert wurden.

Mit Hilfe eines Maßes für die melodische Ähnlichkeit und anhand entsprechender experimenteller Daten ließe sich nun beispielsweise voraussagen, wie ähnlich eine Melodie zum Original ist, die eine Versuchsperson nach x Wiederholungen und nach y Minuten aus dem Gedächtnis wiedergibt oder wiedererkennt.

Erkenntnisse von dieser Qualität sind nicht nur als Elemente der musikpsychologischen Grundlagenforschung interessant, sondern können beispielsweise bei der Parametrisierung von *Music Information Retrieval*-Systemen dazu dienen, Schwellenwerte für Mengen von gefundenen Dokumenten (*retrieved documents*) – d.h. in diesem Kontext Melodien – zu setzen.

Keine oder nur sehr indirekte Aussagen lässt das Konzept der melodischen Ähnlichkeit darüber zu, welche Elemente und Dimensionen einer Melodie genauer oder besser erinnert werden. Die Anzahl der möglichen Melodien, die zu einer Referenzmelodie einen vergleichbaren Ähnlichkeitswert besitzen ist gerade bei längeren Melodien und niedrigeren Ähnlichkeitswerten sehr groß.

In diesem Kapitel soll ein Experiment im Wiedererkennensparadigma vorgestellt werden, das primär dazu dient zu klären, welche Teile und welche Parameter einer Melodie gut bzw. zuerst im Gedächtnis behalten werden und welche anderen Teile oder Parameter später und ungenau. Dabei wird angenommen, dass nicht alle Teile einer Melodie gleich gut oder gleich schnell erinnert werden und sich die Gedächtnisfehler nicht zufällig verteilen, sondern ein mehr oder weniger deutliches Muster zeigen. Von besonderem Interesse ist hier der Bereich der sogenannten ‚unscharfen‘ Gedächtnisrepräsentation, bei dem ein Gedächtnisinhalt, d.h. eine Melodie, weder exakt mit einer externen Melodie übereinstimmt, aber auch nicht vollkommen aus dem Gedächtnis verschwunden ist, sondern z.B. nur in der Akzentstruktur vorhanden ist. Von besonderem Interesse sind die Gedächtnisfehler, die ein unscharfes Erinnern mit sich bringen, und die hier systematisiert und gewichtet werden sollen.

Regelmäßigkeiten in den Gedächtnisrepräsentationen von Melodien können auf mindestens drei verschiedene Arten konzipiert und modelliert werden:

Der erste Modellrahmen verwendet vor allem zeitliche Parameter zur Beschreibung von melodischen Gedächtnisrepräsentationen. Es wäre z.B. denkbar und plausibel, dass die ersten 7 Töne oder der Inhalt der ersten zwei Sekunden einer Melodie besonders genau im Gedächtnis behalten werden können, wenn man von einer begrenzten Kapazität des Kurzzeit- oder Arbeitsgedächtnisses ausgeht und Millers magische Zahl (Miller, 1956) bzw. die zeitliche Spanne der phonologischen Schleifen von Baddeleys *Working Memory* (Baddeley, 1986, 1997) als Grenzen annimmt. Es könnten auch die ersten und die letzten Töne eines Musikstückes oder einer Melodie mit ab- bzw. zunehmender Häufigkeit Gedächtnis behalten werden, wie es nach den in Gedächtnisaufgaben häufig anzutreffenden *primacy*- oder *recency*-Effekten (vgl. Kluwe, 1990, S. 121f) zu erwarten wäre.

Der zweiten Rahmen, in dem sich mögliche Gedächtnismodelle für Melodien bewegen können, wird durch Modelle gebildet, die Gestaltgesetze für die Analyse von melodischen Gebilden verwenden. Dazu gehören das melodisch-rhythmische Konsonanzmodell von Monahan und die *Joint-Accent-Hypothesis* bzw. das Konzept der dynamischen Form von

Boltz und Jones, wie sie im Kapitel 3.3 erläutert wurden. Gemein ist diesen Modellen, dass sie Möglichkeiten eröffnen, die in der Wahrnehmung wichtigeren Töne einer Melodie zu ermitteln. Für diese wichtigeren Töne werden bessere Erinnerungsleistungen angenommen als für die übrigen Töne. Es wird also innerhalb dieses Modellrahmens hypothetisiert, dass der ‚Kern‘ oder die Struktur einer Melodie besser und zuerst erinnert, während es sekundäre Bestandteile einer Melodie gibt, die später und ungenauer erinnert werden. Dies entspricht den Annahmen vieler Analysemethoden und Modelle, die sich der Reduktion von melodischen Gebilden auf ein Melodiegerüst oder eine Melodiestructur bedienen. Diese Vorstellung hat eine lange Geschichte in der Musikwissenschaft, angefangen bei Heinrich Schenkers ‚Urlinie‘ (1932), über die Volksliedforschung (Bartok, 1976; Jesser, 1990) und die generativen Musiktheorien (Sundberg & Lindblom, 1976; Lerdahl & Jackendoff, 1983) bis hin in den Bereich der neueren Gedächtnisforschung, wie sie in Kapitel 3.3 bereits besprochen wurde (Sloboda & Parker, 1985; Oura & Hatano, 1988).

Auch wäre denkbar, dass von einem Musikstück gerade die melodischen Teile besonders gut erinnert werden, die eine wie auch immer zu definierende Prägnanz oder besondere Charakteristik aufweisen. Innerhalb dieser Vorstellung würde neben der Struktur dieser prägnanten Teile auch die konkrete Ausgestaltung eine Rolle spielen.

Der dritte Bereich von Modellen, der auch bereits ausführlich im Kapitel über kurze Tonfolgen erläutert (3.2) wurde, macht dagegen Voraussagen über die musikalischen Dimensionen, innerhalb derer am besten melodische Informationen behalten werden, z.B. melodische Kontur, Intervalle, Rhythmus, harmonische Informationen. Hierhin gehören vor allem die Modelle von Dowling (z.B. 1978), Massaro (Idson & Massaro, 1978; Massaro et al. 1980) und Dewitt & Crowder (1986). Voraussagen ausgehend von diesen Modellen könnten die Genauigkeit der Erinnerung in den verschiedenen Dimensionen einschließen, die gegeneinander gewichtet werden.

Zusätzlich zu diesen drei Modellbereichen, nach denen sich die Eigenschaften des Testmaterials beschreiben lassen, wird auch die bei den Versuchspersonen angesiedelte Größe ‚Musikalische Vorbildung‘ bzw. ‚Beschäftigung mit Musik‘ betrachtet, die in den meisten bisherigen Forschungen ebenfalls als erklärende Größe miteinbezogen wurde.

Obwohl nur sehr wenig Forschungsergebnisse vorliegen, die Voraussagen der drei materialbeschreibenden Modellbereiche miteinander kombinieren, ist es intuitiv sehr plausibel anzunehmen, dass sowohl die Dauer und der zeitliche Abstand, wie auch Faktoren der musikalischen Struktur und die Unterschiede in den musikalischen Dimensionen beim Behal-

ten und Erinnern von Melodien interagieren und zusammen zur Erklärung von Gedächtnisleistungen beitragen. Das Design des hier geschilderten Wiedererkennensexperimentes berücksichtigt diese Möglichkeit zur Interaktion, indem je ein Faktor aus einem Modellbereich als Hauptfaktor aufgenommen wurde. Die Versuchsbedingungen sind unter allen Stufen der drei Hauptfaktoren entsprechend ausbalanciert worden.

Drei verschiedene Ansätze zur Modellbildung über die Experimentaldaten sollen zur Anwendung kommen:

- Der erste Ansatz aggregiert die Antworten der Versuchspersonen mit Maßen aus der *Signal Detection Theory* und vergleicht die Versuchsbedingungen unter den Hauptfaktoren. Dieser Ansatz entspricht dem von Dowling und Kollegen üblicherweise gewählten Vorgehen.
- Der zweite Ansatz verzichtet auf eine Aggregation der Daten und bildet ein Modell mit den Rohdaten der Versuchspersonen, die durch die Hauptfaktoren des Experiments erklärt werden sollen.
- Der dritte Ansatz berücksichtigt neben den Hauptfaktoren alle erhobenen unabhängigen Variablen und sucht mit gängigen Verfahren der statistischen Modellierung oder des sogenannten *data mining* nach einem ‚optimalen‘ Modell zur Erklärung der Rohdaten der Versuchspersonen.

Neben Erkenntnissen über das musikalische Wiedererkennensgedächtnis und den es bedingenden Faktoren wird der Vergleich der drei Ansätze auch darüber Auskunft geben, welche Art von Information aus welchem Ansatz zu gewinnen ist. Der methodische Aspekt ist bei diesem Modellvergleich also von zentraler Bedeutung.

5.2 Ökologische Validität von Design und Versuchsmaterialien

Design und Epistemologie des Experiments entsprechen dem Standard-Paradigma von Wiedererkennensexperimenten, so wie es ausführlich im Kapitel über das Gedächtnis für kurze Tonfolgen (Kapitel 3.2) beschrieben und in zahlreichen Studien zum Gedächtnis für melodischen Materialien verwendet wurde (z.B. Dowling & Fujitani, 1971; Idson & Massaro, 1978; Kauffman & Carlsen 1989; Dowling et al. 2002) wurde. Grundannahme ist, dass die Melodieelemente, die in einer Wiedererkennenaufgabe („Gleich/Verschieden-Bewertung“) häufiger richtig beurteilt werden, besser im Gedächtnis gespeichert sind als die Elemente, die signifikant weniger oft richtig beurteilt werden.

Im Gegensatz zu den älteren Wiedererkennensexperimenten soll dieser Versuch ein größeres Maß an ökologischer Validität durch die Wahl mehrerer versuchstechnischer Optionen und Parameter erhalten. Insgesamt soll eine echte Hörsituation zumindest näherungsweise simuliert werden, bei der eine Person ein Stück Popmusik aufmerksam hört und versucht, sich die Melodien dieses Stückes zu merken. Motivation für das Einprägen der Melodien in das Gedächtnis in solch einer echten Hörsituation könnte z.B. die Möglichkeit sein, die Melodien anschließend in ein *Query-by-humming*-System zu singen oder sie unter mehreren möglichen Melodien aus einer Melodiendatenbank durch Anhören und Vergleich mit dem Gedächtnisinhalt zu identifizieren. Folgende Eigenschaften der Versuchsmaterialien und der Versuchsprozedur sollen zu dieser angestrebten ökologischen Validität beitragen:

- Es werden ‚echte‘, existierende Popstücke verwendet, die in Melodik, Rhythmik und Harmonik der weit verbreiteten Pop- und Schlagermusik entsprechen und den Versuchspersonen deshalb stilistisch vertraut sind.
- Es werden mehrstimmige Klavierarrangements der Popsongs verwendet, die aus Melodie und Begleitung bestehen, sodass Einflüsse beispielsweise auf die Metrum- und Harmoniewahrnehmung, die durch Interaktionen von Melodie- und Begleitstimme beim Hören von Musik in echten Lebenssituationen zustande kommen, berücksichtigt werden.
- Es werden längere Ausschnitte (ca. 50 Sekunden) aus Musikstücken als Präsentationsmaterial und Zielreize (*targets*) zwischen 8 und 18 Sekunden in der Wiedererkennensphase verwendet, sodass Erinnerungsleistungen erforderlich sind, die über die üblicherweise angenommene Spanne des Arbeitsgedächtnisses bzw. der phonologischen Schleife (ca. 2 Sekunden, vgl. Baddeley, 1997) hinausgehen.
- Die Präsentationsausschnitte bestehen jeweils aus mehreren formalen Abschnitten (z.B. Strophe, Refrain, Bridge) und enthalten mehrere verschiedene Melodien. Es werden so retro- und proaktive Interferenzeffekte (vgl. Kluwe, 1990, S. 125) bewusst nicht ausgeschlossen, wie sie auch in echten Hörsituationen als wahrscheinlich angenommen werden können.
- Die verwendeten Popstücke decken eine gewisse Spannweite an musikalischen Stilen (Blues, Rock’n’Roll, Schlager, Popballaden, Blackmusic, Uptempo-Nummern etc.), an Tempi (60 bis 150 Schläge pro Minute) und rhythmischen und harmonischen Gestaltungsmitteln ab. Die verwendeten Melodien passen sich so auf unterschiedliche Weise in die Begleitung ein und zeigen auch eine große Breite an me-

lodischen Formtypen und Gestaltungsmerkmalen. Diese Heterogenität im Untersuchungsmaterial ist mit Absicht gewählt worden, um so die möglicherweise gefundenen Ergebnisse für einen größeren Bereich von Popmusik und den dort verwendeten Melodien generalisieren zu können.

Trotz dieser Maßnahmen bei der Versuchsgestaltung bleiben immer noch eine Reihe von Merkmalen, die das Experiment als Laborversuch kennzeichnen und Unterschiede zu realen Hörsituationen darstellen. Hier sind zu nennen:

- Die Umgebung (Raum mit bürotypischer Ausstattung) ist ein für das Musikhören ungewohnter Ort.
- Die Musik wird im Einzeltest ohne Störungen gehört.
- Die Aufmerksamkeit für die Melodien ist maximal.
- Die Präsentationsstücke sind Klavierarrangements und nicht mit der Originalbesetzung bzw. im Originalsound zu hören.
- Die Gesangsstimme ist durch eine Klaviermelodie ersetzt.
- Es existieren keine gesungenen Texte, die als zusätzliche Gedächtniscues genutzt werden könnten.
- Es werden pro Testsession 24 verschiedene Stücke präsentiert. Interferenzeffekte zwischen den Stücken und Konzentrationseinbußen aufgrund der Versuchsdauer können deshalb nicht ausgeschlossen werden.

Einige dieser Merkmale lassen eine schlechtere, andere eine bessere Gedächtnisleistung als in bestimmten realen Hörsituationen erwarten. Da es jedoch nicht möglich ist, mit vertretbarem finanziellen und zeitlichen Aufwand das Melodiegedächtnis in realen Hörsituationen zu testen, stellen diese Merkmale einen Kompromiss zwischen der notwendigen Homogenisierung der Versuchsbedingungen für alle Versuchsteilnehmer und der möglichst großen Annäherung an die lebensweltliche Wirklichkeit dar, für die das melodische Gedächtnis getestet werden soll.

5.3 Methode

5.3.1 Versuchsteilnehmer

Im Mittelpunkt des Interesses bei diesem Versuch steht das Gedächtnis für Melodien von Erwachsenen, die keine überdurchschnittliche musikalische Vorbildung besitzen, d.h. deren aktive und passive Beschäftigung im Durchschnitt der deutschen Bevölkerung dieser

Altersgruppe entspricht. Aus diesem Grund wurden als Versuchsteilnehmer Musiker, Musik- oder Musikwissenschaftsstudenten im vorhinein ausgeschlossen. Rekrutiert wurden für die insgesamt 42 Versuchsteilnehmer vor allem Studierende anderer Fachrichtungen und berufstätige junge Erwachsene. Das Durchschnittsalter lag bei 29 Jahren (Bereich: 16 bis 63 Jahre; 50% waren zwischen 22 und 33,5 Jahren alt). Die augenblickliche und vergangene aktive sowie passive Beschäftigung mit Musik wurde mittels eines Fragebogens erhoben, der wie der Fragebogen im Experiment zur melodischen Ähnlichkeit am Vorbild von Mainz und Salthouse (1998) angelehnt war und die meisten musikbezogenen Tätigkeiten auf metrischem Skalenniveau erhob.

64,3% der Versuchsteilnehmer spielten ein Instrument oder hatten in der Vergangenheit jemals ein Instrument gespielt oder sich intensiver mit Gesang beschäftigt. Im Durchschnitt hatten die Versuchspersonen, die je aktiv Musik gemacht hatten, ihr Instrument 9,4 Jahre gespielt und 3,3 Jahre bezahlten Instrumentalunterricht genossen. Dies sind deutlich geringere Werte als die, die Musikexperten aus der Musikwissenschaft im ersten Experiment dieser Arbeit bei gleichzeitig viel geringeren Durchschnittsalter (23,2 Jahre) erreicht hatten (12,5 Jahre Instrumentalspiel, 6 Jahre bezahlter Unterricht, vgl. die entsprechenden Daten zu den Versuchspersonen des Ähnlichkeitsexperiments in Kapitel 4).

Im Mittel hörten die Versuchspersonen täglich 3,0 Stunden Musik. All diese Werte können als ‚normal‘ im Vergleich mit der deutschen (erwachsenen) Bevölkerung gelten: Nach der Shell-Studie *Jugend 2000* betätigen sich 23% der deutschen Jugendlichen musikalisch aktiv (Jugendwerk der Deutschen Shell, 2000). Dieser Prozentsatz erhöht sich entsprechend; wenn man die vergangene musikalische Aktivität kumuliert.

Für den Musikkonsum lassen sich die von Rösing zusammengetragenen Angaben als Referenz heranziehen (Rösing, 1998, S. 109). Er gibt einen durchschnittlichen Musikkonsum der Bundesbürger von täglich ca. 3 Stunden an.

5.3.2 Materialien:

5.3.2.1 Die Teststücke:

Als Materialien wurden für das Wiedererkennensexperiment 36 Popstücke ausgewählt. Die Auswahl der Popstücke erfolgte nach folgenden Kriterien:

- Die Stücke sollten dem Kreis der Versuchspersonen möglichst nicht bekannt sein. Die Vorbekanntheit der Stücke wurde zusätzlich während des Experimentes kon-

trolliert und die entsprechenden Beobachtungen bei Vorbekanntheit eines Stückes von einer Vpn. aus der Datenanalyse ausgeschlossen.

- Die Stücke sollten mindestens zwei unterschiedliche, gut sangliche Melodien enthalten, die sich jeweils über mehrere Takte erstrecken. Meist waren dies die Strophen- und Refrainmelodien.
- Die Melodien sollten nicht zu viele Tonrepetitionen enthalten und stattdessen ein gewisses Maß von melodischer und rhythmischer Bewegung zeigen. Rap-Stücke wurden aus diesem Grund nicht ausgewählt.
- Die Stücke sollten einen eingängigen ‚radio-tauglichen‘ Charakter aufweisen, der sie als repräsentativ für die große Menge melodischer Popmusik gelten lassen kann. Dabei wurden Kompositionen älteren Datums (z.B. *Oh Carol* (Neil Sedaka), *Take good care of my baby* (Carole King / Gerry Coffin)) ebenso ausgewählt wie neuere unbekannte Stücke aus den 90er Jahren. Die Auswahl der Stücke erfolgte mit Hilfe von diversen Songbooks (Hit Pack 1-3, AMA-Verlag; Hits der 70er, Hits der 80er & 90er, KDM-Verlag) und aus der umfangreichen *Leadsheet*-Sammlung von Dr. Paul Rigganbach, die vor allem aus GEMA-Anmeldungen in den 90er Jahren für deutsche Musikverlage resultiert. Im Schnitt wurden pro Versuchsperson nur 0,21 Stücke richtig mit Titel erkannt (s. detaillierte Ergebnisse unten), zu denen die Beobachtungen von der Datenanalyse ausgeschlossen werden mussten.

Die Stücke wurden für die Verwendung im Experiment neu arrangiert. Das Arrangement bestand aus einer Klavierbegleitung, die in relativ einfachen, aber gleichzeitig dem Stil der Titel entsprechenden rhythmischen Mustern die zu den Melodien zugehörigen Akkorde präsentierte. Zweck der Begleitung war die möglichst eindeutige Darstellung des harmonischen und metrischen Zusammenhangs der Melodien. Da die Begleitung aber nur den wahrnehmungsmäßigen Hintergrund darstellen sollte, wurde jedem Ton der Begleitung ein MIDI-Lautstärkewert von 50 zugewiesen. Über der Begleitung wurde die Melodiestimme deutlich hervorgehoben in einer Lautstärke von 120 präsentiert. Begleitung und Melodiestimme wurden im MIDI-Sound 1 (Grand Piano) eingespielt. Die Dauer der Stücke wurde auf 48 Sekunden normiert. Die Anzahl der verwendeten Melodien, die Länge von Einleitungs- und Übergangstakten sowie das Tempo der Stücke wurde ggf. angepasst, wobei allzu große Abweichungen vom Originaltempo vermieden wurden. Arrangement und Einspielung der ausgewählten Titel lagen in der Hand von Lars Dammann, der als Keyboardlehrer an der Musikschule Seevetal e.V. große Erfahrung mit dem Arrangieren von Pop-

musikstücken besitzt. Als Beispiel für ein Arrangement der Teststücke ist hier der Titel *Bitter Blue* der deutschen Gruppe Kind of Blue wiedergegeben.

Bitter Blue

Bernd Klimpel

Melodie

Begleitung

1

11

16

21

Abb. 5-1: Arrangement des Titels Bitter Blue (Autor: Bernd Klimpel, Interpreten: Kind of Blue) als Teststück.

5.3.2.2 Die Testmelodien:

Aus den Gesamtarrangements wurden für die Prüfung des Wiedererkennensgedächtnis je eine Test-Melodie isoliert, die den Versuchspersonen einzeln vorgespielt wurde. Die Kriterien für die Auswahl einer bestimmten Melodie als Testmelodie aus einem Stück bezogen sich vor allem auf die Melodielänge und deren Struktur:

- Die Testmelodie sollte nach Möglichkeit ein in sich abgeschlossenes melodisches Gebilde darstellen, also entweder eine ganze Melodie oder zumindest ein eigener Melodieteil, der durch Pausen oder melodische oder harmonische Wendungen einen gewissen Abschluss besitzt.
- Die Testmelodie sollte genügend melodische und rhythmische Bewegung besitzen, um an ihr auf sinnvolle Weise die Manipulationen hinsichtlich Kontur, Intervallen und Rhythmus vornehmen zu können. Insbesondere Tonfolgen mit sehr vielen Tonrepetitionen oder sehr monotonen Rhythmen wurden nicht als Testmelodie gewählt.
- Die Testmelodie sollte mindestens 16 und höchstens 40 Töne besitzen.
- Die Testmelodie sollte im Tempo des zugehörigen Stückes zwischen 8 und 18 Sekunden lang sein.

Voraussetzung für die getrennte Manipulation von Akzenten bzw. Nicht-Akzenten in den Testmelodien war eine Operationalisierung der Akzentregeln nach Monahan und Boltz & Jones für das in diesem Versuch verwendete experimentelle Material. Da die zugehörigen Theorien im Kapitel 3.3 bereits ausführlich dargestellt wurden, soll hier nur der verwendete Algorithmus für die Vergabe der Akzente bzw. Akzentgewichte in Kurzform dargestellt werden. Dabei ist zu beachten, dass die Theorien von Boltz und Jones sowie von Monahan nur sehr vage die Ermittlung von Akzenten in echten Melodien spezifizieren, wie oben bereits angemerkt wurde. Eine ganze Reihe von konkreten Spezifizierungen für die Analyse von Melodien bleibt offen, Parameter, wie z.B. für die Gewichtung der unterschiedlichen Dimensionen werden i.d.R. gar nicht festgelegt. Insofern ist die im folgenden beschriebene Akzentbestimmung für die Konstruktion der Testmelodien eine Heuristik, die sich nicht direkt aus der Literatur ableitet.

Es existieren drei Quellen für Akzente: Die Tönhöhendimension, die rhythmische Dimension und die metrische Dimension. Voraussetzung für die Akzentvergabe ist eine Einteilung der Melodie in Phrasen, die hier nach einem sehr einfachen Verfahren mit Bezug auf die relativen Tondauern vorgenommen wurde:

Phraseneinteilung:

- Es werden nur die Töne ermittelt, mit denen eine Phrase beginnt.
- Eine Phrase dauert vom Onset des ersten Phrasentons (einschließlich) bis zum Onset des Tons, der die nächste Phrase beginnt (ausschließlich).
- Der erste Ton einer Melodie beginnt eine Phrase.
- Es wird die häufigste Tondauer der Melodie ermittelt (Modus).
- Jeder Ton, dem ein Interonset-Intervall (IOI) vorausgeht, das $\geq 4x$ dem Modus der Tondauern ist, beginnt eine Phrase.

Akzentvergabe:

- Gib jedem Ton, der eine Phrase beginnt, einen rhythmischen und einen Tonhöhenakzent.
- Gib jedem Ton, der sein Onset auf den betonten Taktzählzeiten (1 + 3) hat, einen metrischen Akzent.
- Gib jedem Ton, dessen Dauer länger ist als die seines Vorgängertons, einen rhythmischen Akzent.
- Gib jedem Ton, dessen Vorgänger- und Nachfolgeton jeweils höher oder tiefer ist, einen Tonhöhenakzent (Konturumkehrung). Hier wurde keine Bereinigung von Wechselnoten vorgenommen.
- Gib jedem Ton, dem ein Intervall von >3 Halbtöne vorausgeht, einen Tonhöhenakzent (Sprung).

Jeder Ton kann aus einer Quelle nur einen Akzent bekommen. Maximal ist also ein Akzentgewicht von 3 pro Ton möglich.

Da die Verteilung der Akzentgewichte von der konkreten Gestalt der Melodie abhängt, lässt sich nicht im Vorhinein festlegen, ob nun Töne mit einem Akzentgewicht von 3 als die ‚wahren‘ Akzente gelten sollen. Vielmehr erschien es sinnvoll, die ‚wahren‘ Akzente anhand der Verteilung der Akzentgewichte in der Melodie zu definieren, so dass das Verhältnis von akzentuierten zu nicht-akzentuierten Tönen in einem sinnvollen Rahmen liegt. Die Regeln zum Selektieren der wahren Akzente der Testmelodien lauten also:

1. Bestimme die Anzahl der Töne, die zu jeder Akzentgewichtsklasse (0, 1, 2, 3) gehören.
2. Prüfe, ob die Anzahl der Töne mit Akzentgewicht 3 kleiner als 25% aller Töne ist.
3. Falls nein, definiere die Töne der Akzentklasse 3 als ‚wahre‘ Akzente.
4. Prüfe, ob die Anzahl der Töne mit Akzentgewicht 3 größer als 10% aller Töne und kleiner als 33% aller Töne mit Akzentgewicht von mindestens 1 ist.
5. Falls ja, definiere alle Töne der Akzentklasse 3 als ‚wahre‘ Akzente.
6. Falls nein, summiere die Anzahlen der Töne in Akzentklasse 3 und 2.
7. Prüfe, ob die Anzahl der Töne mit Akzentgewichten 3 oder 2 größer als 10% und kleiner als 25% aller Töne und kleiner als 33% aller Töne mit Akzentgewicht von mindestens 1 ist.
8. Falls ja, definiere alle Töne der Akzentklassen 3 und 2 als ‚wahre‘ Akzente.

Eine weitere Schleife mit der Summierung der Töne Akzentklasse 1 zu den Klassen 2 und 3 ist theoretisch denkbar, war in der Stichprobe der ausgewählten Testmelodien jedoch an keiner Stelle notwendig.

Zur Veranschaulichung der Phraseneinteilung und Akzentbestimmung ist hier die aus *Bitter Blue* entnommene Testmelodie (dritte und letzte Melodie des Teststücks). Ein senkrechter Strich markiert jeweils einen Phrasenbeginn, die waagerechten Pfeile kennzeichnen die Töne, die nach dem Akzentvergabemechanismus als ‚wahre‘ Akzente gelten.



Abb. 5-2: Akzente und Phrasensegmentierung in der Testmelodie aus dem Teststück *Bitter Blue*.

Die 36 mit Akzenten versehenen Testmelodien wurden in drei Gruppen hinsichtlich der Hypothesen eingeteilt: Die Testmelodien der Gruppe IDT waren identische Kopien einer Melodie aus dem zugehörigen Teststück. In Gruppe MNA (Nicht-Akzent manipuliert) befanden sich Melodien, die sich in einigen Tönen oder Tondauern von der entsprechenden Melodie im jeweiligen Teststück unterschieden. Die unterschiedliche Töne oder Tondauern betrafen jedoch nicht solche, die nach den Akzentregeln von Monahan und Boltz & Jones

als Akzente klassifiziert worden wären. Die Gruppe MAK (Akzent manipuliert) schließlich enthielt Melodien, bei denen Töne und Tondauern mit Akzent manipuliert worden waren. Ein Vortest mit insgesamt 5 Versuchspersonen hatte ergeben, dass in diesem experimentellen *setting* sinnvollerweise 4 Töne bzw. Tondauern zu manipulieren sind, um für Versuchspersonen ohne spezielle musikalische Vorbildung einen geeigneten Schwierigkeitsgrad zu erreichen, bei dem es weder zu *ceiling*- und noch *floor*-Effekten kommt; d.h. die Versuchspersonen sollten Ergebnisse erreichen, die weder zu dicht an der Ratewahrscheinlichkeit noch an einer perfekten Diskriminierung der Stimuli liegen.

Die Manipulation der Töne wurde je in einer der drei musikalischen Dimensionen Rhythmus, Intervalle und Kontur vorgenommen.

Bei Rhythmusmanipulationen wurden vier einzelne Töne verlängert oder verkürzt. Da dies in der Regel zwangsläufig auch Auswirkungen auf die Dauer und den Rhythmus der benachbarten Töne hatte, konnten maximal also 8 Töne in ihrer Dauer verändert sein. Die Definition einer Änderung lehnt sich hier an das Konzept der Edit Distance an: Wenn durch das Austauschen nur eines Dauernwertes eine Übereinstimmung zwischen Original und manipulierter Variante hergestellt werden kann, so liegt genau eine Manipulation vor. In der Gruppe der akzentmanipulierten Melodien (MAK) wurden Töne mit Akzent in ihrer Dauer verändert. Bei den nicht-akzentmanipulierten Melodien (MNA) wurden die Akzente soweit wie möglich in ihrer Dauer belassen.

Die Intervallmanipulationen veränderten das Intervall eines Tones zum Vorgänger und/oder Nachfolgeton, wobei die Intervallrichtung (Kontur) erhalten blieb. Auch hier wurden jeweils vier Töne verändert. Der manipulierte Ton wurde jeweils so verändert, dass auch nach der Änderung eine musikalisch sinnvolle Melodielinie erhalten bleibt, d.h. es wurden keine tonartfremden Töne eingesetzt, es wurden große Sprünge vermieden und die minimalste Intervalländerung bevorzugt. Um der harmonischen Dimension Rechnung zu tragen, die vor allem nach den Untersuchungen von Boltz (1993) für das Erkennen von Unterschieden zwischen Melodien von Bedeutung ist, wurden die manipulierten Töne so gewählt, dass sie nicht in dem gleichzeitig erklingenden Akkord (in der Begleitung) enthalten waren. Diese Bedingungen waren für die Melodien der Gruppen MNA und MAK identisch.

Die Konturmanipulationen waren Intervallveränderungen, bei denen gleichzeitig, die Intervallrichtung verändert wurde. Auch hier wurden vier Töne pro Testmelodie verändert.

Es galten die selben Regeln für die Manipulation der Töne wie bei den Intervallmanipulationen.

Die intervallmanipulierte Testmelodie von *Bitter Blue* ist hier dargestellt, um die Art der Manipulationen besser nachvollziehen zu können. Sie enthält gegenüber der Originalmelodie (s.o.) vier Intervallmanipulationen auf akzentuierten Tönen (Takt 2, Takt 4 und 2x in Takt 5).



Abb. 5-3: Intervallmanipulierte Testmelodie aus dem Teststück *Bitter Blue*

5.3.3 Prozedur

Die Aufgabe der Versuchspersonen bestand darin, je ein 48-sekündiges Teststück und danach die zugehörige Testmelodie aufmerksam anzuhören und anschließend anzugeben, ob die Testmelodie identisch und unverändert in dem vorangegangenen Teststück vorhanden war oder ob die Testmelodie im Vergleich mit der Referenzmelodie im Teststück irgendwie verändert war. Den Versuchspersonen stand ein Fragebogen mit einer sechsstufigen Skala zur Verfügung, die neben dem Identitätsurteil auch die subjektive Urteilssicherheit berücksichtigte („sehr sicher identisch“, „sicher identisch“, „identisch“, „verschieden“, „sicher verschieden“, „sehr sicher verschieden“). Da es sich bei den Teststücken um echte und zum Teil auch veröffentlichte Popsongs handelt, wurde mit zwei weiteren Fragen (*items*) des Fragebogens die mögliche Vorbekanntheit der Teststücke abgefragt. Zum einen stand den Versuchspersonen eine fünfstufige Skala zur Verfügung, auf der sie die Sicherheit angeben sollten, mit der ihnen das Teststück bekannt vorkam („sicher nein“, „vielleicht nein“, „weiß nicht“, „vielleicht ja“, „sicher ja“). Zum anderen wurden die Versuchspersonen gebeten, zu jedem bekannten Stück Titel und Interpret anzugeben.

Jede Vpn. hatte 24 Paare von Teststück und Melodie zu bewerten. Teststück und zugehörige Melodie waren jeweils durch eine Pause von 2 Sekunden getrennt, zwischen zwei Paaren gab eine Pause von 8 Sekunden den Versuchspersonen Zeit, die beiden *items* auf dem Fragebogen anzukreuzen. Nach zwölf Paaren wurde eine Pause von ca. 3 Minuten einge-

legt, um eine Abnahme der Konzentration in der zweiten Testhälfte zu vermeiden. In dieser Pause wurden die Versuchspersonen informell nach der von ihnen verwendeten Strategie bei der Bewältigung der Aufgabe befragt. Die Versuchspersonen wurden einzeln getestet. Die Teststücke und Melodien wurden ihnen über Kopfhörer direkt von einem CD-Spieler vorgespielt.

Dem eigentlichen Experiment vorangestellt war eine Trainingsphase. Anhand von drei Beispielen wurden jeder Versuchsperson der Ablauf und die Art des Testmaterials vorgeführt. Nach jedem Trainingsbeispiel bekamen die Versuchspersonen die richtige Antwort („identisch“ bzw. „verschieden“) als Feedback mitgeteilt.

Nach der Experimentalphase hatten die Versuchspersonen noch einen Fragebogen zu ihren aktiven und passiven Musikgewohnheiten in Gegenwart und Zukunft auszufüllen. Der Fragebogen glich dem für das Ähnlichkeitsexperiment verwendeten (s. Kap. 4) und orientierte sich an dem von Meinz & Salthouse (1998) verwendeten.

Die Dauer des eigentlichen Experiments betrug ca. 25 Minuten, inklusive Versuchsinstruktionen, Trainingsphase und Ausfüllen des Fragebogens belief sich eine Session auf ca. 50 Minuten.

5.3.4 Design und Hauptfaktoren

Da sich in Vortests gezeigt hatte, dass die Konzentration der Versuchspersonen bei dieser Art von Experiment nach mehr als 30 Minuten deutlich nachlässt bzw. Verwechslungen zwischen den einzelnen Testmelodien häufiger werden und so unerwünschte Positionseffekte der Testabfolge auftreten, wurde die Anzahl der Teststücke pro Vpn. auf 24 begrenzt. Zudem wurden die Versuchspersonen in drei gleichstarke Gruppen zu je 14 eingeteilt und mit einer verschiedenen Auswahl von 24 aus der Gesamtheit der 36 *Testitems* (Paare von Teststücken und Melodien) getestet. Die Auswahl der *items* für die Gruppen geschah natürlich so, dass jedes *item* insgesamt gleich häufig (28 Mal) beurteilt wurde. Die Reihenfolge der *items* innerhalb der drei Testgruppen wurde durch eine Zufallsverteilung realisiert.

Die drei Hauptfaktoren, unter denen die Bedingungen ausbalanciert wurden, waren die folgenden:

- *Position der Referenzmelodie*: 1. Faktorstufe. im 1. Teil des Stückes (Melodieende der Referenzmelodie ist mehr als 20 Sekunden vom Stückende entfernt); 2. Stufe:

im 2. Teil des Stückes (Melodieanfang ist nicht mehr als 20 Sekunden vom Stückende entfernt).

- *Klasse der Manipulation*: 1. Faktorstufe: keine (IDT); 2. Stufe: Manipulation, Akzenttöne intakt (MNA); 3. Stufe: Akzente manipuliert (MAK).
- *Art der Manipulation*: 1. Faktorstufe: keine (wenn Klasse der Manipulation = IDT), 2. Stufe Rhythmus; 3. Stufe: Intervalle; 4. Stufe: Kontur.

Wie leicht nachvollziehbar ist, kann in keiner der drei Dimensionen eine Manipulation auftreten, wenn die Test- und Referenzmelodie identisch sind. Das bedeutet, dass die drei Faktoren nicht vollständig von einander unabhängig sind. Während die Faktoren 1 und 2 vollkommen unabhängig von einander sind, hängt Faktor 3 und Faktor 2 ab, d.h. nur unter der 2. und 3. Stufe des 2. Faktors können sich die 2., 3. und 4. Stufe des 3. Faktors finden und unter der 1. Stufe des 2. Faktors kann sich nur die erste Stufe des 3. Faktors finden.

Durch diese inhaltlich bedingte Hierarchie der Faktoren 2 und 3 ist das Design im eigentlichen Sinne nicht vollständig ausbalanciert, sondern stellt ein quasi-ausbalanciertes Design dar. Statt der theoretisch 24 Bedingungen (2x3x4) reduziert sich das Design auf 14 Bedingungen. Dies veranschaulicht folgende Tabelle, bei der jeder Zeile eine Bedingung entspricht:

Bedingung	Position Referenzmelodie	Manipulationsklasse	Manipulationstyp
1	1	1	1
2	1	2	2
3	1	2	3
4	1	2	4
5	1	3	2
6	1	3	3
7	1	3	4
8	2	1	1
9	2	2	2
10	2	2	3
11	2	2	4
12	2	3	2
13	2	3	3
14	2	3	4

Tabelle 5-1: Die 14 Experimentalbedingungen mit den zugehörigen Ausprägungen der drei Faktoren des Designs.

5.3.5 Weitere Einflussgrößen und Operationalisierungen

Insgesamt ergibt das hier verwendete Versuchsdesign maximal 1008 Beobachtungen (42 Versuchspersonen x 24 Melodien). Allerdings lieferten einige wenige Versuchspersonen fehlende Werte für einzelne *items*, und es wurden die Werte der *items* entfernt, die der jeweiligen Vpn. mit Sicherheit vorbekannt waren (s.u.). Insgesamt blieben 995 Beobachtungswerte in der Auswertung.

Als abhängige Variablen wurden für die Berechnungen verschiedene Maße eingesetzt bzw. miteinander verglichen. Neben den Rohwerten aus der sechsstufigen Beurteilungsskala, wurden die je vier Werte jeder Vpn. unter jeder Faktorstufenkombination der Faktoren *Position der Referenzmelodie* und *Klasse der Manipulation* aggregiert; und zwar sowohl als Mittelwerte wie auch in Anlehnung an das Vorgehen von Dowling et al. (2002) als Mediane und als *area scores* (AUC) unter den *receiver-operating*-Kurven (ROC). Die ROCs und die zugehörigen AUCs wurden jeweils im Vergleich der Werte der nicht-manipulierten Melodien mit den akzent-manipulierten bzw. den nicht-akzent-manipulierten Melodien für beide Stufen des Faktors *Position der Referenzmelodie* getrennt berechnet. Eine weitere Aggregation der Werte pro Vpn. unter dem zusätzlichen Faktor *Art der Manipulation* schien nicht sinnvoll, da hier für eine Vpn. pro Stufe nur ein bzw. zwei Werte vorlagen.

Ohne Berücksichtigung der Urteilssicherheit wurden zusätzlich die Antworten der Versuchspersonen auf die beiden Kategorien falsch und richtig für jedes *item* reduziert. Als abhängige Variablen ließen sich hieraus demnach die Anzahl der richtigen Antworten (pro Versuchsbedingung) und als *bias*-bereinigtes Maß d' für den Vergleich der manipulierten Bedingungen mit den identischen Bedingungen berechnen. Auch d' wurde von Dowling und seinen Kollegen bereits häufig in ähnlichen Situationen verwendet (z.B. Dowling, 1973b, 1986, 1991).

Die latente Einflussgröße *Musikalische Vorbildung* musste aus den Angaben aus dem Fragebogen durch Datenverdichtung ermittelt bzw. operationalisiert werden. Bei den Musikexperten des ersten Experiments (Kap. 4) erfolgte dies über eine Faktorenanalyse, die die metrisch skalierten Variablen zum aktiven Musizieren auf drei Dimensionen abbildete. Da in der Stichprobe dieses Experiments jedoch mehr als ein Drittel der Versuchspersonen nie ein Instrument gespielt hat, ist diese Methode aufgrund der vielen fehlenden Werte schon

aus formalen Gründen nicht sinnvoll. Zusätzlich ist die musikalische Vorbildung bei Laien mit Sicherheit auch inhaltlich anders zu bestimmen. Denn hier dürfte eine wichtige Dimension durch die Ausprägung passiver Musikgewohnheiten, wie Häufigkeit und Intensität des Musikhörens und anderer musik-orientierter Tätigkeiten, gegeben sein. Der Fragebogen widmete 8 *items* auf dichotomem, ordinalen und metrischen Niveau den passiven Musikgewohnheiten.

Zusätzlich zu den drei Hauptfaktoren *Position der Referenzmelodie*, *Klasse der Manipulation*, *Art der Manipulation* und der *Musikalischen Vorbildung* wurden noch 13 weitere Größen der getesteten Musikstücke bzw. Melodien berücksichtigt, die für das Wiedererkennen der Testmelodie potentiell verantwortlich sein könnten. Diese Variablen gehören zum ersten bzw. dritten in der Einleitung dieses Kapitels aufgespannten Modellrahmen. Es handelt sich um Variablen, die zum einen zeitliche Zusammenhänge zwischen dem Auftauchen der Referenzmelodie im Teststück und der Testmelodie messen; zum andern sind es Eigenschaften auf verschiedenen (einfachen) musikalischen Dimensionen, die zwischen den beiden Melodien verglichen werden.

Für diese Variablen lag jedoch keine systematische Variation innerhalb der Versuchsmaterialien vor, sondern ihre Ausprägungen wurden als zufällig verteilte Ausprägungen bzw. zufällig verteilte Werte einer Kovariablen festgehalten. Zusammen mit den Hauptfaktoren und der musikalischen Vorbildung sind die Werte dieser 13 Variablen das Datenmaterial, anhand dessen in einer *data-mining*-Analyse ein möglichst genaues Modell der für das Melodiegedächtnis relevanten Faktoren und ihres jeweiligen Gewichtes ermittelt werden soll. Die 13 Variablen gliedern sich in drei Gruppen, gemäß ihrer inhaltlichen Ausrichtung. Die ersten Variablen messen die Position und die Stärke der Manipulationen. Die zweite Gruppe gibt Unterschiede in der melodischen Struktur zwischen den beiden zu vergleichenden Melodien, die in der obigen Operationalisierung der Struktur nicht berücksichtigt worden waren. Die dritte Gruppe misst den Umfang des musikalischen Geschehens zwischen der Referenzmelodie im Teststück und der Vergleichsmelodie bzw. den Umfang der Vergleichsmelodie selbst:

Gruppe 1

- *Mittlere Position der Manipulationen*: Es wird der Abstand der manipulierten Töne in Sekunden vom Ende des Teststücks bestimmt und das arithmetische Mittel be-

rechnet. Nach den Angaben in der Literatur sollte mit zunehmender Zeit zwischen den manipulierten Tönen und der Vergleichsmelodie die Gedächtnisleistung abnehmen. Diese Variable misst also im Prinzip das gleiche wie der Faktor *Position der Referenzmelodie* jedoch auf metrischem Niveau und in Sekunden.

- *Manipulation von Note in Phrase*: Reihenfolgezahl der ersten manipulierten Note innerhalb ihrer Phrase; wenn die manipulierten Noten alle aus der selben Phrase stammten, wurde das arithmetische Mittel ihrer Reihenfolgezahlen bestimmt. Diese Variable misst also die Position einer Manipulation. Allerdings wird eher von einem *primacy*-Effekt ausgegangen. Demnach sollten sich die ersten Noten einer Phrase besser in das Gedächtnis einprägen als die Noten im weiteren Verlauf einer Phrase.
- *Manipulation von Phrase in Melodie*: Reihenfolgezahl der ersten Phrase mit einem manipulierten Ton im Teststück. Auch hier wird eher ein *primacy*-Effekt erwartet, d.h. besonders die erste Phrase einer Melodie sollte gut im Gedächtnis bleiben.
- *Akzentgewicht der Manipulation*: Es wird das Akzentgewicht der manipulierten Töne nach oben erläuteter Methode ermittelt und das arithmetischen Mittel berechnet.

Gruppe 2

- *Note Teil von repetierter Struktur*: Binär kodierte Variable, die angibt, ob mindestens eine Note Teil einer exakten Wiederholung innerhalb der Testmelodie ist. Die dahinterstehende Annahme geht davon aus, dass bei Wiederholungen innerhalb der Testmelodie, bei denen ein Teil manipuliert ist, der andere jedoch nicht, die Versuchspersonen aufgrund des Vergleichs innerhalb der Testmelodie urteilen (können).
- *Änderung der harmonischen Funktion*: Obwohl die Manipulationen sich immer im Rahmen der Diatonik bewegten, kann der manipulierte Ton den harmonischen Kontext ändern. Dies dürfte ggf. leichter von den Versuchspersonen wiederzuerkennen sein als die eigentliche Kontur- oder Intervalländerung. Die Änderung der harmonischen Funktion ist folgendermaßen definiert: Ein Melodieton kann in dem gleichzeitig erklingenden Akkord des Teststückes enthalten sein oder nicht und bekommt einen binären Wert der Akkordzugehörigkeit (0,1) zugeteilt. Eine harmonische Änderung liegt dann vor, wenn sich die Zugehörigkeit mindestens eines lan-

gen Tones (länger als Vorgänger- und Nachfolger) oder eines Tones auf den Takt-schwerpunkten (1 und 3 im 4/4-Takt) zwischen manipulierter und nicht-manipulierter Testmelodie geändert hat.

- *Änderung des Ambitus*: Potentiell leicht zu entdecken ist die Veränderung des Ambitus einer Melodie, denn hierfür ist im vorliegenden Design ein tonales Gedächtnis für absolute Tonhöhen ausreichend. Die Variable misst in binärer Form, ob die manipulierte Testmelodie den höchsten bzw. tiefsten Ton der nicht-manipulierten Melodie über- oder unterschreitet.
- *Melodische Ähnlichkeit*: Die Ähnlichkeit zwischen Referenzmelodie im Stück und der Testmelodie in der Wiedererkennensphase, gemessen mit dem in Kapitel entwickelten Maß *optil*, das auf den Vergleich von melodischen Varianten der selben Ursprungsmelodie optimiert wurde.

Gruppe 3

- *Anzahl der Noten zwischen Ende der Referenzmelodie und Stückeende*: Anzunehmen ist, dass die Töne, die zwischen dem Ende der Referenzmelodie im Teststück und der Vergleichsmelodie zu hören sind, negative Interferenzeffekte auslösen oder die Inhalte in einem begrenzten Gedächtnissystem ‚überschreiben‘.
- *Anzahl der Melodien zwischen Ende der Referenzmelodie und Stückeende*: Geht von der selben Annahme aus wie die vorangehende Variable, misst aber in Melodieeinheiten.
- *Dauer zwischen Ende der Referenzmelodie und Stückeende*: Selbe Annahme wie die beide vorangegangenen Variablen, aber in Sekunden gemessen.
- *Dauer der Testmelodie*: gemessen in Sekunden. Eine längere Testmelodie müsste – bei konstanter Zahl der Manipulationen – mehr Material zur Wiedererkennung liefern und deshalb auch höhere Werte auf der *rating*-Skala (in Richtung des identischen Pols) generieren.
- *Länge der Testmelodie*: Selbe Annahme wie vorige Variable, jedoch in Anzahl der Noten gemessen.

5.4 Datenauswertung und erste Ergebnisse

Vor den eigentlichen Ansätzen zur Modellbildung ist es sinnvoll, sich einen Überblick über die durch den Fragebogen erhobenen Daten zu verschaffen. Von Bedeutung sind hier die *ratings*, mit denen die Versuchspersonen angaben, mit welcher Sicherheit ihnen ein Teststück bekannt war, die verwendeten Gedächtnisstrategien, die richtigen Identifizierungen einzelner Teststücke und eine kurze soziodemographische Beschreibung der Stichprobe. Etwas ausführlicher werden danach die musikalische Vorbildung der Versuchsteilnehmer untersucht, die gemäß der Hypothesen einen Einfluss auf die Gedächtnisleistungen haben soll, sowie die Verteilung der Gedächtnisleistungen selbst.

5.4.1 Richtige Identifizierung einzelner Teststücke

Wie oben geschildert wurde insgesamt neunmal ein Teststück von einer Vpn. richtig mit dem Namen des Interpreten oder des Titels identifiziert. Dies sind 0,89% der 1008 Beobachtungswerte. Da maximal drei Versuchspersonen die selbe Melodie erkannt hatten, bestand kein Anlass, einzelne Melodien aufgrund ihrer Vorbekanntheit in der Stichprobe der Versuchspersonen auszuschließen. Erkannt wurden von den Versuchspersonen „Take good care of my baby“ (2x) von Carole King, „While my guitar gently weeps“ von den Beatles (3x), „Oh Carol“ von Neil Sedaka (2x), „Cold, cold heart“ von Marti Pellow (1x) und „Bitter Blue“ von der deutschen Band Kind of Blue (1x). Diese Stücke stammten ausnahmslos aus den Hit-Songbooks. Keines der Stücke aus der Sammlung von Dr. Riggensbach, die wie gesagt in der Mehrzahl unveröffentlicht sind, wurde identifiziert.

Die *rating*-Werte zu den neun Fällen, in denen ein Stück richtig identifiziert wurde, wurden von den nachfolgenden Datenanalysen ausgeschlossen. Da sich zusätzlich noch vier fehlende Werte bei den *rating*-Daten ergeben hatten, blieben insgesamt 995 Beobachtungen in der Analyse.

5.4.2 Subjektive Bekanntheit des Testmaterials

In den neun Fällen, bei denen eine Vpn. das Teststück richtig identifizieren konnte, gab die jeweilige Vpn. mit einer Sicherheit von 5 (höchster Wert) an, das Teststück zu kennen. In diesen Fällen waren also die Bewertungen der Bekanntheit korrekt. Interessanter ist jedoch die Frage, inwiefern auch hohe Bekanntheitsbewertungen ohne Identifizierung des Teststückes darauf schließen lassen, dass der Vpn. das Teststück vorbekannt war und sie deshalb einen Vorteil bei der Bewältigung der Gedächtnisaufgabe besaß. Hier lässt sich an die Un-

tersuchung von Peynircioğlu et al. (1998) anknüpfen, die den sogenannten „feeling of knowing“-Zustand (vgl. Hart, 1965) für Melodien und Titel von Musikstücken untersuchte.

Bei einem Mittelwertsvergleich über einen t-Test wurden die Mittelwerte auf der Bekanntheitsskala für die veröffentlichten und die nicht-veröffentlichten Stücke verglichen (unter Ausschluss der Fälle mit richtiger Identifikation). Hier zeigte sich eine im Mittel leicht aber signifikant höhere Bekanntheit für die veröffentlichten Teststücke (arithmetisches Mittel(*publish*): 2,51; *aM(non-publish)*: 2,17; T: -3,736, df: 792, $p < 0,000$).

Auch eine Überprüfung des Zusammenhangs zwischen der Vergabe des Wertes ‚5‘ auf der Bekanntheitsskala und dem dichotomen Merkmal der Veröffentlichung der Teststücke in einem χ^2 -Test ergab ein signifikantes Ergebnis (χ^2 -Wert nach Pearson: 9,769, $p = 0,001$).

Diese Werte sprechen dafür, dass die Vpn. die Teststücke, denen sie einen hohen Bekanntheitswert bzw. eine 5 vergeben haben, wirklich erkannt haben können, allerdings weder Interpret noch Titel angeben konnten. Dieses Ergebnis entspricht der alltäglichen Erfahrung, dass das Gedächtnis für Melodien bzw. erklingende Musik nicht mit dem Gedächtnis für sprachliche Label für Musik kongruent ist.

Angesichts dieser Ergebnisse lässt sich die Frage stellen, ob Fälle, in denen ein hoher Bekanntheitswert vergeben wurde, generell aus der Datenanalyse ausgeschlossen werden sollten, da hier die Versuchspersonen einen Vorteil durch die Vorbekanntheit der Melodie besaßen. Die Ergebnisse der Studie von Peynircioğlu et al. (1998) sprechen jedoch eher gegen dieses Vorgehen. Denn Peynircioğlu et al. fanden bei ihren Versuchspersonen generell höhere Bekanntheitswerte für Melodien als für Musiktitel bei bekannten Stücken. Das Gefühl eine Melodie wiederzuerkennen, stellt sich also offenbar schnell ein, wenn zumindest kleinere Teile aus der Melodie wiedererkannt werden. Einen echten Gedächtnisvorteil zeigten in jener Studie jedoch nur die Versuchspersonen, die auch den Titel der Melodie richtig erinnerten: Melodien, die von den Versuchspersonen nach einem gegebenen Anfang richtig weitergesungen wurden, wurden gleichzeitig auch besser durch den Titel erinnert bzw. wiedererkannt, wenn der Titel vorgegeben wurde. D.h. wenn die Versuchspersonen ein nachweisbar besseres Gedächtnis für die Melodie eines Stückes besaßen, erinnerten sie auch signifikant öfter den Titel. Umgekehrt scheint ein Wiedererkennen der Melodie ohne Erinnern des Titels für eine ungenaue oder unvollständige Gedächtnisrepräsentation der Melodie zu sprechen. Deshalb wurde entschieden, die Fälle, in denen die Vpn.

eine hohen Bekanntheitswert vergeben hatte, das Teststück aber nicht mit Interpret oder Titel identifizieren konnten, in der Datenanalyse zu belassen.

5.4.3 Verwendete Gedächtnisstrategien

Nach der Hälfte der zu bewertenden *items*, zwischen den Teilen der Experimentalsitzung, wurde eine kurze Pause von ca. drei Minuten eingelegt, um die Konzentrationsleistung der Versuchspersonen über die gesamte Sitzung zu erhalten und die Erinnerung an die vorangegangenen *Testitems* aus dem Gedächtnis zu löschen, um so Interferenzeffekte zu vermeiden.

Inhaltlich wurde diese kurze Pause durch eine Befragung der Vpn. gefüllt, die auf die Ermittlung der Strategie zur Bewältigung der Testaufgabe abzielte. Neben der direkten Frage, wie eine Vpn. sich die Melodien eines Stückes zu merken versuchte, wurde auch der musikalische Hintergrund und nach spezifischen Kenntnisse (z.B. Notenlesen) gefragt. Die Antworten wurden nicht systematisch, beispielsweise über einen Fragebogen, aufgenommen und es wurde vermieden, den Versuchspersonen andere mögliche Strategien zur Bewältigung des Versuches aufzuzeigen oder vorzuschlagen, um die Testergebnisse nicht zu beeinflussen. Tatsächlich sollte den Versuchspersonen der Eindruck einer Pause vermittelt werden, die durch ein lockeres Gespräch gefüllt wird. Eine echte Interviewsituation wäre hierfür kontraproduktiv gewesen.

Auch ohne systematische Aufzeichnung in der Testphase ließen sich im Nachhinein die verwendeten Strategien notieren und sehr leicht in wenige unterschiedliche Vorgehensweisen einordnen. Die folgenden sechs Strategien sind als qualitative Ergebnisse dieser Befragung zu verstehen. Eine Auszählung von Häufigkeiten oder eine feste Kodierung der Strategien mit Zuordnung zu einzelnen Versuchspersonen wäre in diesem experimentellen Design nicht sinnvoll durchführbar gewesen. Denn die Versuchspersonen waren frei in der Wahl und im Wechsel ihrer Gedächtnisstrategien. Tatsächlich berichteten einige, dass sie je nach Teststück die eine oder die andere Strategie verwendeten, je nachdem, welche ihnen individuell für das bestimmte Teststück als einfacher oder effizienter erschien. Darüber hinaus verwendeten viele Versuchspersonen eine Kombination von Strategien bei der Bewältigung des selben *Testitems*. So waren z.B. die Assoziation von Melodieteilen zu anderen bekannten Melodien und die Prüfung der Plausibilität der Testmelodie (s.u.) gut kombinierbare Strategien. Um die Effizienz unterschiedlicher Gedächtnisstrategien hinsichtlich der Testaufgabe zu beurteilen, ist diese Auswertung also nicht geeignet. Hier hätte bei-

spielsweise analog dem Design von Davis und Yelland (1977) die Vorgabe der Gedächtnisstrategie Teil des experimentellen Designs sein müssen.

Die folgenden sechs Strategien können nach dem Ergebnis der Befragungen unterschieden werden:

Übertragung der melodischen Kontur in die visuelle Vorstellung: Gerade die musikalisch Aktiveren versuchten häufig über eine Rekodierung der melodischen Kontur in ein visuelles ‚Auf und Ab‘, die Erinnerung an die verschiedenen Melodien eines Teststückes zu erhalten bzw. zu präzisieren. Einige Versuchspersonen, die die Notenschrift beherrschten, nahmen die Distanz der imaginierten Töne auf dem Notenpapier als Vorstellungshilfe.

Inneres Nachsingen oder Wiederholung im Gedächtnis: Die häufigste Technik war das lautlose singende Wiederholen von Melodieteilen oder Motiven oder das nochmalige Durchlaufen des melodischen Materials im Gedächtnis. Schwierigkeiten bereitete hier den Versuchspersonen, dass die Teststücke zwei oder mehr verschiedene Melodien enthielten und es den meisten, die diese Strategie verwendeten, sehr schwer viel, mehrere Melodien gleichzeitig oder nacheinander zu wiederholen. Eine Wiederholung der jeweils letzten Melodie ist bei vielen Versuchspersonen deshalb häufiger aufgetreten.

Speicherung des Rhythmus: Einige Versuchspersonen gaben an, sich hauptsächlich am Rhythmus der Melodie oder einzelner Motive zu orientieren, da dieser für sie häufig prägnanter bzw. besser zu erinnern sei. Als Gedächtnisstütze wurde dabei häufig das Mitklopfen der Taktschläge mit dem Fuß verwendet, welches z.B. eine Erschließung von Synkopen ermöglicht.

Plausibilität der Testmelodie: Einige Versuchspersonen erklärten, dass ihnen das aktive Einprägen der Melodien ins Gedächtnis sehr schwer fiel. Dadurch, dass immer mehrere Melodien zu behalten seien und gegenseitig Störbedingungen füreinander darstellten, sei ihnen das bewusste Einprägen fast unmöglich. Aus diesem Grund verlegten sich einige Versuchspersonen darauf, das Teststück zwar aufmerksam zu hören, selbst aber erst in der Wiedererkennungphase am Ende aktiv zu werden. Hier entschieden sie anhand der Plausibilität der Testmelodie, d.h. anhand der Sinnfälligkeit der Intervallverbindungen und des Rhythmus der einstimmigen Melodien. Sie ließen sich dabei i.d.R. von der Annahme leiten, dass wenig sinnfällige oder ungewöhnliche Ton- oder Dauernfolgen eher auf eine manipulierte, also nicht identische Testmelodie hinweisen. Eine in den ‚Ohren‘ der Vpn. sehr

sinnfällige Melodie wäre demnach eher als identisch zur Referenzmelodie im Teststück zu bewerten.

Wiedererkennen von spezifischen Merkmalen: Sehr nahe verwandt mit der letztgenannten Strategie ist das aufmerksame Hören und das spätere Entdecken von kurzen spezifischen Eigenheiten einer Melodie. Auch hier verzichteten die Versuchspersonen bewusst auf eine aktive Enkodierung während des Hörens des Teststückes, hofften aber spezielle oder außergewöhnliche Eigenschaften, wie ungewöhnliche Intervalle oder einen besonderen Rhythmus in der Testmelodie wiederzuerkennen, der auch zuvor in der Referenzmelodie präsent war. Die Gefahr, Unterschiede zwischen Referenz- und Testmelodie bei dieser Gedächtnisstrategie zu übersehen, ist natürlich gegeben.

Assoziation zu bekannten Melodien: Nicht wenige Versuchspersonen nutzten ihr Assoziationsgedächtnis, um spezielle Motive oder Intervallwendungen mit ihnen bekannten Melodien in Verbindung zu bringen. Es konnte so eine Rekodierung des melodischen Materials stattfinden. Dabei war es für die Vpn. unerheblich, ob der *Assoziationscues* in der Referenzmelodie wirklich dem assoziierten Inhalt sehr ähnlich war oder nicht. Ausreichend ist die Möglichkeit, individuell eine Verbindung herzustellen, die hinterher erinnerbar ist. Die Testmelodie wurde dann mit dem assoziierten Inhalt bzw. mit der Referenzmelodie als Abweichung vom assoziierten Inhalt im Gedächtnis verglichen.

Diese qualitative Einteilung und Beschreibung der unterschiedlichen Strategien zur Bewältigung der Versuchsaufgabe kann zwar nicht unmittelbar mit den numerischen Ergebnissen der Versuchsauswertung verrechnet werden, die im weiteren Verlauf dieses Kapitels generiert werden. Allerdings können die Erkenntnisse über die verwendeten Gedächtnisstrategien wertvolle Hinweise bei der Interpretation der statistischen Ergebnisse liefern und sind für die weitere Modellbildung von zentraler Bedeutung. Denn erst wenn bekannt oder zumindest erahnbar ist, welche Informationen die Versuchspersonen bei der Bewältigung der Aufgabe dieses Experimentes genutzt haben, ist es auch möglich, diese statistisch zu modellieren.

5.4.4 Soziodemographische Beschreibung der Stichprobe

Da der soziodemographische Hintergrund der Versuchsteilnehmer nicht in Beziehung zu deren Gedächtnisleistung gesetzt werden sollte, beschränkte sich die Erhebung desselben auf Alter, Geschlecht und höchster absolvierter Bildungsabschluss. In Hinsicht auf diese

Merkmale sollte die Stichprobe zumindest in etwa vergleichbar sein mit einer potentiellen Nutzerschaft von *Music-Information-Retrieval*-Systemen, wie dem für diese Arbeit anvisierten *Query-by-humming*-System.

Die Versuchsteilnehmer waren im Mittel 29 Jahre alt (Standardabweichung: 8,7 Jahre). Dies lässt sich als relativ typisch für musikinteressierte Internetnutzer verbuchen, wenn man als Vergleich die Netobserver-Erhebung, eine große Nutzerbefragung im Internet vom April 2003 durch das Marktforschungsinstitut Novatris, heranzieht: Dort war das Durchschnittsalter der Nutzer von Musikseiten 27,5 Jahre (Novatris, 2003).

64,3% der Versuchspersonen waren weiblich und 35,7% männlich. Hier ist der Unterschied deutlich zu den üblichen Geschlechterquoten der Internetnutzung.

Die Häufigkeitsverteilung zum höchsten erreichten Bildungsgrad lässt erkennen, dass es sich bei den Versuchspersonen vorwiegend um Studierende handelt, die ihr Studium noch nicht abgeschlossen haben: 50% der Befragten besaßen ein Abiturzeugnis, weitere 28% sogar ein abgeschlossenes Studium, wie der nächsten Grafik zu entnehmen ist.

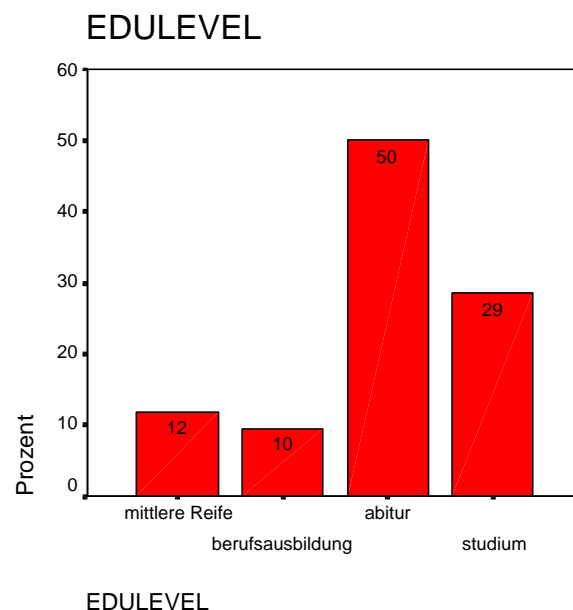


Abb. 5-4: Verteilung höchsten erreichten Bildungsabschlüsse in der Stichprobe der 42 Versuchspersonen in Prozent.

Da Studenten generell als sehr internetaffine Personengruppe gelten, spricht auch dieser Befund für die Eignung der Stichprobe.

Abgesehen von dem für die Internetnutzung untypischen Geschlechterverhältnis lässt sich die Stichprobe der Versuchsteilnehmer als in etwa vergleichbar mit der Zielgruppe von Internet-basierten *Music-Information-Retrieval*-Systemen bezeichnen.

5.4.5 Die musikalische Vorbildung der Versuchsteilnehmer

Im Gegensatz zum soziodemographischen Hintergrund wurde für die musikalische Vorbildung der Versuchsteilnehmer ein deutlicher Einfluss auf deren melodische Gedächtnisfähigkeiten angenommen. Aus diesem Grund waren 17 *items* des Fragebogens der Erhebung der musikalischen Vorbildung gewidmet.

Die folgende Tabelle gibt die deskriptiven Statistiken zu den ordinalen und metrischen Variablen des Fragebogens wieder. Auch hier zeigt sich wieder eine große Nähe zu den musikinteressierten Internetnutzern der Novatris-Studie. Die deskriptiven Ergebnisse zu den aktiven und passiven musikalischen Gewohnheiten der Versuchsteilnehmer sind in der folgenden Tabelle zusammengefasst.

Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung	Varianz
Musikhören/Tag	42	,25	12,00	3,0298	2,68021	7,184
CDSBUY	42	,00	5,00	,8929	1,26653	1,604
CDSBURN	42	,00	10,00	2,2857	2,54027	6,453
CONCERTS	42	,00	5,00	,6905	1,29705	1,682
MUSICMAG	42	0	1	,07	,261	,068
SINGALON	41	1	4	1,90	,831	,690
YEARSINS	26	1,50	30,00	9,3654	7,07751	50,091
PRACTMOM	27	,00	3,00	,4074	,79707	,635
MUSMAKMO	27	,00	7,00	,9444	1,76686	3,122
PRACTPAS	27	,00	20,00	4,5741	4,03995	16,321
MUSMAKPA	27	,50	40,00	6,1667	7,44725	55,462
PAIDLESS	27	,00	156,00	39,3333	42,33202	1792,000
PAIDGIGS	27	,00	5,00	,4815	1,25178	1,567
GIGS	26	,00	40,00	11,5577	12,86416	165,487
Gültige Werte (Listenweise)	24					

Tabelle 5-2: Deskriptive Kennzahlen zum musikalischen Hintergrund der Versuchsteilnehmer. (CDSBUY = Anzahl der im letzten Monat gekauften CDs; CDSBURN = Anzahl der im letzten Monat gebrannten CDs; CONCERTS = Anzahl der im letzten Monat besuchten Konzerte; MUSICMAG = Anzahl der im letzten Monat gekauften Musikzeitschriften; SINGALON = Häufigkeit des Singens für sich alleine; YEARSINS: Anzahl Jahre, in denen ein Instrument gespielt wurde; PRACTMOM = Stunden pro Woche, die z.Z. geübt wird; MUSMAKMO = Stunden pro Woche, in denen z.Z. Musik gemacht wird; PRACTPAS = Stunden pro Woche, die in der musikalisch aktivsten Phase geübt wurde; MUSMAKPA = = Stunden pro Woche, in denen in der musikalisch aktivsten Phase Musik gemacht wurde; PAIDLESS = Anzahl der Monate, in denen für Musikunterricht gezahlt wurde; PAIDGIGS = Anzahl der bezahlten absolvierten Auftritte; GIGS = Anzahl der absolvierten Auftritte).

Die Variable *singalon* gibt die Antworten der Versuchspersonen auf die Frage „Singst Du für Dich alleine“ auf ordinalem Niveau wieder, wobei 1 für „oft“ steht und 4 für „nie“. Zusätzlich wurde auf nominalem Niveau noch nach dem Chorsingen gefragt, dem gegenwärtigen oder vergangenen Instrumentalspiel und nach den liebsten Freizeitbeschäftigungen.

gen. In einem Chor sangen nur 3% der Versuchspersonen, dagegen spielten 64% ein Instrument oder hatten in der Vergangenheit schon einmal ein Instrument gespielt. Schließlich zählten 52,5% eine musikalische Tätigkeit (wie Musikhören oder Musikmachen) zu ihren drei liebsten Hobbys.

Weitere interessante Erkenntnisse ließen sich aus den erhobenen Daten gewinnen, wie z.B. die im Monatsmittel deutlich höhere Zahl gebrannter oder geschenkter CDs (2,3) gegenüber den gekauften CDs (0,9), die die Versuchspersonen angaben anzufertigen bzw. zu erwerben. Jedoch soll hier lediglich festgehalten werden, dass die Stichprobe für die Population der musikinteressierten (jungen) Erwachsenen insgesamt als typisch gelten kann.

5.4.5.1 Indexbildung und Datenverdichtung anhand der Variablen zur musikalischen Vorbildung

Die Variablen mussten zur Beantwortung der Frage, ob die musikalische Vorbildung einen Einfluss auf die Leistungen des Melodiegedächtnisses ausübt, auf wenige Dimensionen verdichtet werden. Da sich die erhobenen Merkmale im Gegensatz zum Fragebogen für die Musikwissenschaftsstudenten im ersten Experiment dieser Arbeit nicht ausschließlich auf metrischem Niveau messen ließen, wurde die Indexbildung als Verfahren zur Datenverdichtung der metrischen, ordinalen und nominalen Variablen gewählt.

Eine anhand der existierenden Literatur (z.B. Bruhn & Rösing, 1993; Rösing & Bruhn, 1993; Rösing, 1998) plausible a-priori-Einteilung der Versuchspersonen nach ihrer musikalischen Vorbildung, reduziert die erhobenen Variablen auf die Quantität des Musikkonsums (d.h. der passiven Beschäftigung mit Musik) und des aktiven Musikmachens. Die Indexvariablen für beide Dimensionen, die sich jeweils aus der Verrechnung mehrerer Merkmale des Fragebogens ergeben, haben so zunächst ein metrisches Skalenniveau. Um für die Varianzanalysen, die der Beantwortung der Haupthypothesen dienen, Faktoren mit einer geringen Zahl an Stufen zu gewinnen, ist es dann nötig, die metrischen Indexvariablen in wenige, d.h. hier in je drei Kategorien einzuteilen. Die beiden Dimensionen werden hier mit *Musikkonsum* und *Musikalische Aktivität* bezeichnet. Als Ausprägungen werden jeweils die Kategorien ‚niedrig‘, ‚mittel‘ und ‚hoch‘ unterschieden. Die Dimensionen werden zunächst als unabhängig voneinander angenommen. Eine ‚Verrechnung‘ der beiden Dimensionen, d.h. eine Verdichtung auf eine einzige Dimension, ist deshalb (a priori) nicht sinnvoll.

Die Konstruktion der Indexwerte erfolgte nach inhaltlichen Kriterien und wurde für beide Dimensionen getrennt in folgenden Schritten vorgenommen:

- Einteilung der insgesamt 17 Variablen zum musikalischen Hintergrund in Variablen zum Musikkonsum und zur Musikalischen Aktivität (s.u.).
- z-Standardisierung der Variablen
- Berechnung eines gewichteten Mittels für die beiden Dimensionen. Dabei wurden diejenigen Variablen höher gewichtet, die einen größeren Einsatz von persönlichen Ressourcen (Zeit oder Geld) von der einzelnen Person verlangen, um der musikrelevanten Tätigkeit nachzugehen. Zusätzlich wurden auch die Tätigkeiten hoch bewertet, die potentiell der Person einen Vorteil bei der Bewältigung der Gedächtnis-Testaufgabe verschaffen. Aus diesen Gründen ist beispielsweise die Anzahl der gekauften CDs pro Monat höher gewichtet worden als die der gebrannten CDs bzw. die Gesangsaktivitäten höher als die Jahre des Instrumentalspiels. Sehr wichtig wurde für den Musikkonsum erachtet, ob die Person eine musikalische Tätigkeit zu ihren Lieblingshobbys zählt. Für die musikalische Aktivität war das Chorsingen von großer Bedeutung, da hier in der Regel die Verarbeitung und Reproduktion von Melodien besonders trainiert wird. Die Gewichtung der einzelnen Variablen gibt folgende Tabelle wieder:

Variablen Musikkonsum	Gewichtung	Variablen Musikalische Aktivität	Gewichtung
Musik als Hobby	3	Chorsingen	3
Musikhören / Tag (in Stunden)	1	Häufigkeit alleine singen	2
Gekaufte CDs / Monat	2	Instrumentalspiel	2
Gebrannte CDs / Monat	1	Jahre des Instrumentalspiels	1
Konzertbesuche	2	Üben zur Zeit	1
Musikzeitschrift gekauft	2	Musikmachen zur Zeit	1
		Üben in aktivster Phase	2
		Musikmachen in aktivster Phase	2
		Zeitraum bezahlter Musikunterricht	1
		Anzahl bezahlter Auftritte	3
		Anzahl Auftritte	2

Tabelle 5-3: Gewichtung der verschiedenen Variablen zu den Musikgewohnheiten für die Indexbildung.

- z-Standardisierung der Indexvariablen.

- Einteilung der Versuchsteilnehmer auf den beiden Dimensionen in geringe (30% Quantil), mittlere (Quantil von 31% bis 70%) und hohe Werte (Quantil von 71% bis 100%) anhand der Werteverteilung.

Im Ergebnis wurde jeder Versuchsteilnehmer eindeutig je einer der drei Stufen jeder Dimension zugeordnet.

Dass die beiden Dimensionen des musikalischen Hintergrunds, wie sie hier konstruiert wurden, doch nicht ganz unabhängig voneinander sind, zeigt eine Betrachtung ihrer Korrelationen: Zwar ergab sich für die Pearson-Bravais-Korrelation nur ein nicht signifikanter Zusammenhang von 0,199 zwischen den metrischen, z-standardisierten Indexvariablen, der nicht-parametrische Zusammenhang nach Spearman war allerdings signifikant ($p: 0,425$; $p < 0,0025$ bei einseitiger Testung). Auch für den Zusammenhang zwischen den auf drei Stufen reduzierten Werten der beiden Dimensionen zeigte *Spearman's ρ* einen signifikanten Wert ($p: 0,375$, $p < 0,007$). Da der Zusammenhang zwischen den beiden Dimensionen *Musikkonsum* und *musikalische Aktivität* jedoch nur sehr moderat ausfällt – in jedem Fall $< 0,43$ – wurden beide Dimensionen als getrennte Variablen beibehalten.

5.4.6 Gedächtnisleistungen der Versuchsteilnehmer

Vor der eigentlichen Überprüfung der Haupthypothesen scheint es sinnvoll, die Leistungen der Versuchsteilnehmer im Gedächtnisexperiment näher zu betrachten. Als erste Variable ist hier die Summe der richtigen Antworten von Interesse. Dazu wurde die sechsstufige Beurteilungsskala auf zwei Kategorien (identisch / verschieden) und die Antworten der Vpn. mit den richtigen Lösungen verglichen. Da jede Vpn. 24 *items* zu beurteilen hatte, konnten maximal 24 richtige Antworten erreicht werden. Die Verteilung der richtigen Antworten veranschaulicht folgende Grafik.

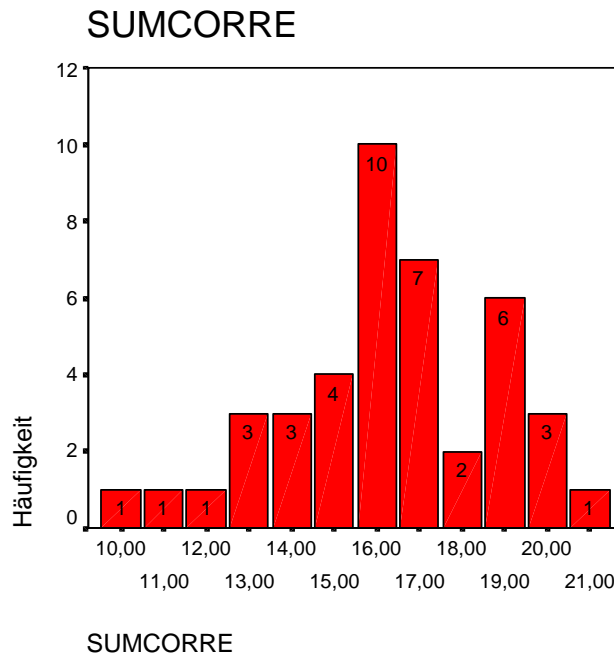


Abb. 5-5: Häufigkeitsverteilung der Summe der richtigen Antworten (SUMCORRE) im Gedächtnistest.

Nach dem Anpassungstest, der auf der z-Statistik nach Kolmogorov-Smirnov beruht, ist die Verteilung einer Normalverteilung hinreichend ähnlich (z: 0,941, exakte Signifikanz: 0,326). Der Mittelwert der Stichprobe liegt bei 16,3, Median ist 16. Die Streuung hält sich in Grenzen (Standardabweichung: 2,5; Varianz: 6,3).

Es stellt sich die Frage, bei welchen Versuchspersonen man ausschließen kann, dass sie nur durch Raten dieses Ergebnis erzielt haben. Die Ratewahrscheinlichkeit liegt im Experiment bei 0,5, die Anzahl der *items* (Versuche) ist 24. Man kann also mit einer Irrtumswahrscheinlichkeit von höchstens 0,05 ausschließen, dass die Versuchspersonen, die 16 oder mehr richtige Antworten gegeben haben, diese durch Raten erzielen konnten, wie die entsprechende Binomialverteilung (vgl. Formel in Abschnitt 4.3.1.2) angibt. Das bedeutet, dass nur 13 Versuchspersonen (31%) unterhalb dieser 95%igen Sicherheit liegen.

Die Ergebnisse der Stichprobe scheinen also gut geeignet zu sein, um die Hypothesen beantworten zu können, denn bei der großen Mehrzahl der Versuchspersonen kann man davon ausgehen, dass sie die gestellte Aufgabe erfüllen konnten. Für die Gesamtheit der Ergebnisse ist weder ein *floor*-Effekt (nahe dem Erwartungswert durch Raten von 12), noch ein *ceiling*-Effekt (nahe dem Maximalwert von 24) festzustellen.

Um einen ersten Eindruck vom Zusammenhang zwischen Testergebnis und den im letzten Abschnitt konstruierten Variablen des musikalischen Hintergrunds zu erhalten, wurden die Korrelationen zwischen Testergebnis (Summe korrekter Antworten) und den beiden Di-

mensionen zum musikspezifischen Hintergrund errechnet. Dies wurde jeweils für die z-standardisierten Indexvariablen sowie für die in drei Stufen eingeteilten Werte berechnet, und zwar jeweils als parametrische Pearson- und als nicht-parametrischen Rangkorrelation nach Spearman. Von den insgesamt acht Korrelationen blieb bei einseitiger Testung nur eine (mit $p < 0,073$) knapp über dem üblichen Niveau der Irrtumswahrscheinlichkeit. Diese nicht-signifikante Korrelation ergab sich für den Zusammenhang zwischen der metrischen Variante der musikalischen Aktivität und der Anzahl korrekter Antworten.

Auch beim Vergleich der durchschnittlichen Testergebnisse zwischen den Versuchspersonen mit geringer und hoher musikalischer Aktivität bzw. geringem und hohem Musikkonsum ergaben sich signifikante Unterschiede in den entsprechenden t-Tests:

Musikalische Aktivität (gering vs. hoch): $T: -2,315, df: 22, p < 0,015$ (einseitig);

Musikkonsum (gering vs. hoch): $T: -2,016, df: 22, p < 0,028$ (einseitig);

Obwohl sich unter Einbeziehung der mittleren Kategorie beider Dimensionen weniger signifikante Ergebnisse in den entsprechenden Tests fanden, lässt sich insgesamt festhalten, dass der musikalische Hintergrund einen positiven Einfluss auf die Testergebnisse hatte. Die Konstruktion der Indexvariablen ist also in dem Sinne erfolgreich gewesen, als dass dieser Zusammenhang, welcher sich aus der bisherigen Literatur (z.B. Monahan et al., 1987; Oura & Hatano, 1988; Kauffman & Carlsen, 1989; Smith & Cuddy, 1989) erwarten ließ, nachgewiesen werden konnte.

5.4.7 Unterschiede zwischen den Melodien

Ein weiterer wichtiger Aspekt, der Hinweise über Einflussfaktoren auf die Daten geben kann, die durch die Faktoren des Designs nicht ausbalanciert wurden, sind die Unterschiede zwischen den einzelnen Testmelodien. Da jedoch das verwendete Design jede Melodie nur einmal in einer gewissen Faktorstufenkombination benutzte, ist es nur indirekt möglich, die Gedächtnisleistung für die einzelnen Melodien ohne den Einfluss der Manipulationen zu evaluieren, indem man z.B. die Unterschiede zwischen den Gedächtnisleistungen für die *items* mit identischer Testmelodie betrachtet.

Ein χ^2 -Test nach Pearson ergab signifikante Unterschiede sowohl zwischen den Gedächtnisleistungen (korrekte Antworten) für alle 36 unterschiedlichen items ($\chi^2: 127,9; df: 35; p < 0,001$) als auch zwischen den zwölf mit identischer Testmelodie ($\chi^2: 35,91; df: 11; p < 0,001$).

Betrachtet man die zugehörige Kreuztabelle im Anhang, die die richtigen und falschen Antworten zu den einzelnen Melodien aufführen, so fallen die gewaltigen Unterschiede bei der Beurteilung zwischen einzelnen Melodien auf. Während bei dem Song *Give me a sign* (Nr. 3) von J. Streifling von 68% der Versuchspersonen die Testmelodie fälschlich als ‚verschieden‘ bezeichnet wurde, identifizierten 86% die Testmelodie von *Innocence Kids* (Nr. 5) von D. Skoupy richtigerweise als ‚identisch‘ zu einem Ausschnitt des Teststücks.

Beide Testmelodien stammten aus unveröffentlichten Stücken, beide waren identisch zu dem Ausschnitt aus dem jeweiligen Teststück, beide Testmelodien waren dem Anfang des Stückes entnommen und beide Stücke gehören einem ähnlichen Stil an (Pop-/Rock-Ballade).

Das Ergebnis des χ^2 -Tests und das Beispiel von *Give me a sign* und *Innocence Kids* deuten darauf hin, dass es Eigenschaften der einzelnen Melodien bzw. Stücke selbst geben muss, die für das Wiedererkennen bedeutsam sind. Dies wird weiter unten bei der statistischen Modellierung der Experimentaldaten zu berücksichtigen sein.

5.5 Statistische Modellierung

Wie zu Beginn dieses Kapitels erläutert sollen die Daten dieses Experimentes mit vier verschiedenen Ansätzen ausgewertet werden. Ein Vergleich der Ergebnisse aus diesen vier Ansätzen soll nicht nur Auskunft über die das Melodiegedächtnis bedingenden Variablen geben, sondern auch aufzeigen, welche Art von Information mit welchem Ansatz zu gewinnen ist, sodass durch die Diskussion der drei Ansätze auch ein methodologischer Erkenntnisgewinn zu verzeichnen ist. Denn wie zu zeigen sein wird, bedingt die Wahl des Ansatzes ganz erheblich den Aspekt des psychischen Gedächtnisprozesses, der modelliert werden soll bzw. die Perspektive auf das, was ein Gedächtnisprozess ist.

Zunächst folgt eine Auswertung nach dem von Dowling und anderen häufig verwendeten Schema, bei dem die Antworten der Versuchspersonen mit Maßen aus der Signal Detection Theory aggregiert und dann varianzanalytisch unter den Hauptfaktoren des Versuchsdesigns verglichen werden.

Danach folgt ein Ansatz, der auf eine Aggregation der Daten verzichtet, aber auch nur die in der Ausbalancierung der Versuchsmaterialien eingesetzten Faktoren verwendet.

Dritte und vierter Auswertungsansatz sind unter dem Schlagwort *data mining* zusammenzufassen. Sie berücksichtigen alle erhobenen unabhängigen Variablen und versuchen, durch Variablenselektion und anschließender Parameterschätzung bzw. Auffinden von

Unterscheidungsmerkmalen die nicht-aggregierten Antworten der Versuchspersonen möglichst genau vorherzusagen. Der dritte Ansatz bedient sich dabei eines linearen Modells, der vierte der sogenannten Klassifizierungs- und Regressionsbäume.

5.5.1 Datenaggregation mit Maßen aus der *Signal Detection Theory*

Dowling verwendet in den meisten seiner Studien zum Wiedererkennensgedächtnis (z.B. Dowling, 1973b; Dowling, 1986; Dowling, 1991; Dowling et al., 1995; Dowling et al., 2002) für Melodien eines der beiden Maße – d' oder AUC (MOC) – aus der Signal Detection Theory als abhängige Variable. Wie im Unterkapitel 3.2 dargestellt, werden zur Berechnung dieser Maße immer die Antworten zweier Versuchsbedingungen miteinander verglichen und daraus ein Differenzwert berechnet. Das AUC-Maß berücksichtigt zusätzlich zur Korrektheit der Antwort noch die Urteilssicherheit auf einer *rating*-Skala.

Da AUC und d' die korrekten Antworten der jeweiligen Versuchsbedingung mit je derselben Referenzbedingung miteinander vergleichen, messen sie in der Tat die Leistung der Versuchspersonen unter einer gewissen Bedingung im Test. In diesen Maßen ist die objektive Übereinstimmung zwischen der Melodie im Teststück und der Vergleichsmelodie der Maßstab, an dem die Antworten der Versuchspersonen gemessen werden. Andere Faktoren, die für die Wahrnehmung und Erinnerung von Melodien eine Rolle spielen können, werden innerhalb einer Versuchsbedingung nicht betrachtet.

Da mehrere Antworten einer Vpn. unter einer Bedingung aggregiert werden müssen, kamen für dieses Modell nur die beiden Faktoren *Position der Referenzmelodie* und *Manipulationsklasse* in Frage. Es ergab sich unter diesen Faktoren folgende Matrix der sechs Versuchsbedingungen:

	Position Referenzmelodie (1. Teil)	Position Referenzmelodie (2. Teil)
Manipulationsklasse (identisch)	1	4
Manipulationsklasse (nicht akzentmanipuliert)	2	5
Manipulationsklasse (ak- zentmanipuliert)	3	6

Tabelle 5-4: Die sechs Versuchsbedingungen unter den Faktorstufen der Hauptfaktoren *Manipulationsklasse* und *Position Referenzmelodie*.

Zu jeder dieser sechs Versuchsbedingungen lieferte jede Vpn. 4 Werte. Eine weitere Unterteilung unter Berücksichtigung des vierstufigen Faktors Manipulationstyp wäre nicht durchführbar gewesen, da hier jede Vpn. pro Bedingung nur einen Wert geliefert hätte.

Für die Berechnung von AUC und d' wurden jeweils die manipulierten Bedingungen mit der identischen verglichen und zwar für die Positionen der Testmelodie im ersten und im zweiten Teil getrennt. Es ergaben sich also sowohl für AUC als auch für d' vier Bedingungen unter zwei Faktoren für die Varianzanalyse

	Position Referenzmelodie (1. Teil)	Position Referenzmelodie (2. Teil)
Manipulationsklasse (iden- tisch vs. nicht akzentmani- puliert)	1	3
Manipulationsklasse (iden- tisch vs. akzentmanipuliert)	2	4

Tabelle 5-5: Die vier möglichen Versuchsbedingungen für die Analyse mit den Maßen d' und AUC unter den Faktorstufen der Hauptfaktoren *Manipulationsklasse* und *Position Referenzmelodie*.

Zusätzlich wurde noch als dritter Faktor ganz analog zu den Auswertungen von Dowling die musikalische Vorbildung im Sinne von musikalischer Aktivität berücksichtigt. Hier wurde die post-hoc vorgenommene und gerade beschriebene Einteilung in drei Niveaus verwendet. Das endgültige Design dieses Modells ist also ein ANOVA-Modell mit 2 x 2 x 3 Faktorstufen, die Anzahl der Beobachtungen für diese Modell war $n = 168$. Ein Levene-Test auf Gleichheit der Fehlervarianzen zeigte keine signifikanten Unterschiede in den Fehlervarianzen über die Gruppen des Designs an. Es gab also keinen Hinweis darauf, dass das dreifaktorielle Modell nicht zulässig wäre.

Als Ergebnis zeigt sich, dass zwar die Unterschiede zwischen den Faktorstufen aller Faktoren in der erwarteten Richtung liegen, dass aber weder die Haupteffekte noch die Interaktionseffekte im Modell signifikant wurden:

- Die Gedächtnisleistung war besser für die Melodien, die aus dem ersten Teil des Teststückes stammen: Mittel der AUC-Werte 1. Teil: 0,741, 2. Teil: 0,701;
- Die akzentmanipulierten Melodien wurden besser von den identischen unterschieden (Mittelwert: 0,73) als die nicht-akzentmanipulierten (0,712);
- Die Versuchspersonen mit größerer musikalischer Aktivität erzielten höhere AUC-Werte als die mit geringerer Aktivität im Musikbereich (gering: 0,69, mittel: 0,703, hoch: 0,77).

Der Faktor *Musikalische Aktivität* war denn auch der einzige, der mit einem p-Werte von 0,078 in die Nähe der üblichen Irrtumswahrscheinlichkeit gelangte. Die beiden anderen Faktoren sowie die Interaktionseffekte waren nicht signifikant, wie die folgende Tabelle zeigt:

Tests der Zwischensubjekteffekte

Abhängige Variable: AUCSCORE

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	,458 ^a	11	,042	1,194	,295
Konstanter Term	84,185	1	84,185	2415,416	,000
POCOMEL	,066	1	,066	1,908	,169
MANCOMP	,013	1	,013	,361	,549
MUSICTACT	,181	2	,091	2,597	,078
POCOMEL * MANCOMP	2,812E-07	1	2,812E-07	,000	,998
POCOMEL * MUSICTACT	,080	2	,040	1,146	,321
MANCOMP * MUSICTACT	,081	2	,041	1,169	,313
POCOMEL * MANCOMP * MUSICTACT	,048	2	,024	,695	,501
Fehler	5,437	156	,035		
Gesamt	92,581	168			
Korrigierte Gesamtvariation	5,895	167			

a. R-Quadrat = ,078 (korrigiertes R-Quadrat = ,013)

Tabelle 5-6: ANOVA-Modell mit den AUC-Werten der Versuchspersonen als abhängiger Variable. (POCOMEL = Position der Referenzmelodie; MANCOMP = Manipulationsklasse; MUSICTACT = Musikalische Aktivität)

Auch R^2 als Kriterium der Modellgüte zeigte mit eine sehr geringe Varianzaufklärung an ($R^2 = 0,078$, korrigiertes $R^2 = 0,013$). Insgesamt erklärt dieses Modell mit den AUC-Werte als abhängige Variable die Gedächtnisleistungen der Versuchspersonen relativ schlecht.

Ein ähnliches Bild ergibt sich bei der Verwendung von d' als abhängige Variable. Bis auf den Faktor *Manipulationsklasse* gehen alle Mittelwertsunterschiede in die erwartete Richtung, bis auf die musikalische Aktivität kommt jedoch weder ein Haupteffekt noch ein Interaktionseffekt in die Nähe des üblichen Signifikanzniveaus von 95%. Dies verdeutlicht die folgende Tabelle:

Tests der Zwischensubjekteffekte

Abhängige Variable: D

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	5,485 ^a	11	,499	,868	,573
Konstanter Term	135,174	1	135,174	235,385	,000
MANCOMP	,002	1	,002	,004	,952
POCOMEL	,778	1	,778	1,354	,246
MUSICTACT	2,829	2	1,414	2,463	,089
MANCOMP * POCOMEL	,056	1	,056	,098	,754
MANCOMP * MUSICTACT	1,508	2	,754	1,313	,272
POCOMEL * MUSICTACT	,033	2	,017	,029	,971
MANCOMP * POCOMEL * MUSICTACT	,298	2	,149	,260	,772
Fehler	89,585	156	,574		
Gesamt	233,026	168			
Korrigierte Gesamtvariation	95,071	167			

a. R-Quadrat = ,058 (korrigiertes R-Quadrat = -,009)

Tabelle 5-7: ANOVA-Modell mit den d' -Werten der Versuchspersonen als abhängiger Variable. (POCOMEL = Position der Referenzmelodie; MANCOMP = Manipulationsklasse; MUSICTACT = Musikalische Aktivität)

Auch mit d' als abhängiger Variable werden die Gedächtnisleistungen von dem dreifaktoriellen Modell sehr schlecht erklärt, wie die Werte für R^2 (0,058) und das korrigierte R^2 (0,009) angeben.

Das dreifaktorielle Design mit *Position der Referenzmelodie*, *Manipulationsklasse* und *Musikalischer Aktivität* der Versuchspersonen ist also ungeeignet, um die Gedächtnisleistungen, die mit Hilfe der Signal-Detection-Maße AUC und d' erfasst werden, zu erklären.

Wie die folgende Grafik der *hits* und *false alarms* (vgl. Erklärungen in Kap. 3.2) in den sechs unterschiedlichen Versuchsbedingungen zeigt, steigt die Erkennungsrate der identischen Melodien (*hits*) bei Melodien aus dem 2. Teil des Teststücks an (Bedingung 4), gleichzeitig schnellen aber auch die *false alarms* bei den manipulierten Melodien in die

Höhe. Genau den gleichen Effekt beobachteten Dowling und Mitarbeiter (2002) bei ihren Experimenten (Exp. 1 und 2) mit unterschiedlichen Retentionsintervallen.

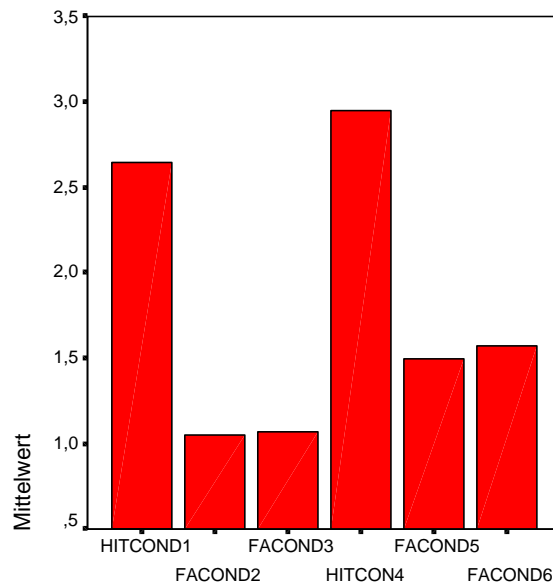


Abb. 5-6: *Hits* und *false alarms* zu den sechs unterschiedlichen Versuchsbedingungen (die Variablenbezeichnungen sind generisch: HIT = *hits*, FA = *false alarms*; COND = *Versuchsbedingung*).

Die Position der Melodie im Teststück wirkt sich weder positiv noch negativ auf die Gedächtnisleistungen der Versuchspersonen aus.

Wie ist über diesen zeitlichen Effekt hinaus das Ausbleiben von signifikanten Effekten gegenüber den relativ analogen Experimenten von Dowling et al. (2002) zu erklären? Da bei dieser Art von psychologischen Experimenten immer mit einer Anzahl von Störgrößen, nicht erklärbarem ‚Rauschen‘ in den Daten und Zufällen zu rechnen ist, kann das Ausbleiben von Effekten in einer Replikation eines Experimentes nicht inhaltlich nachgewiesen oder zweifelsfrei begründet werden. Denn es ist auch möglich, dass sich bei Dowling (innerhalb der Grenzen der angegebenen Irrtumswahrscheinlichkeiten) signifikante Effekte zufällig ergeben haben oder sich diese Effekte in dieser Replikation zufällig nicht ergeben haben.

Trotzdem erscheint es hier sinnvoll, auf einige Unterschiede zwischen den Experimenten von Dowling und Kollegen von 2002 und dem hier beschriebenen Versuch einzugehen, um deutlich zu machen, da es sich hier nicht um eine exakte Replikation des früheren Experimentes handelt, sondern das (möglicherweise substantielle) Unterschiede zwischen den experimentellen Ansätzen bestehen.

Der Fokus dieses Experiments im Gegensatz zu dem von Dowling und Kollegen besteht in der Konzentration auf die ungenauen oder unscharfen Gedächtnisrepräsentationen von melodischem Material. Aus diesem Grund wurden die Unterschiede in den manipulierten Testmelodien sehr gering gehalten und keine vollkommen verschiedenen *Distraktors* in das Design aufgenommen, so dass feine Unterscheidungen zur Bedeutung einzelner musikalischer Parameter oder einzelner struktureller Eigenschaften der Melodien getroffen werden konnten. Referenzmelodie und manipulierte Testmelodie unterschieden sich jeweils in nur vier Tönen, wobei die Testmelodie weder transponiert noch im Tempo verändert war. Zudem mussten die Versuchspersonen die Melodie aus dem musikalischen Ganzen von Melodie und Begleitung, wie es im Teststück zu hören war, abstrahieren.

Die Unterschiede zwischen identischen und ähnlichen *Testitems* in den insgesamt fünf Experimenten von Dowling und Kollegen waren dagegen von viel größerer Natur:

Identische und ähnliche *items* glichen sich in der melodischen Kontur und dem rhythmischen Grundmuster, unterschieden sich in der Tonlage, der Textur (Anzahl der gleichzeitiger Stimmen, Ambitus und Dichte), den Begleitakkorden bzw. dem Begleitrhythmus oder einer Kombination dieser Eigenschaften.

Weiterhin verwendete Dowling neben den ähnlichen *items* echte Distraktoren, also *Testitems*, die vollkommen unähnlich zu den bisher gehörten Phrasen waren und sich somit äußerst stark von möglichen identischen oder ähnlichen *items* unterschieden. Dass die Unterscheidungsfähigkeit von identischen/ähnlichen zu identischen/verschiedenen *items* ansteigt, ist somit zu erwarten.

Im Versuch von Dowling und Mitarbeitern wurde das *Testitem* der ersten oder der zweiten Phrase des Teststückes entnommen und wurde innerhalb der ersten zehn Sekunden präsentiert. Dies stellt eine deutliche Erleichterung für die Versuchspersonen dar, da sie nur die ersten zwei Phrasen für den späteren Vergleich im Gedächtnis behalten mussten.

Die *Testitems* bei Dowling und Kollegen waren mit maximal 6,7 Sekunden deutlich kürzer als die hier präsentierten Testmelodien (8-20 Sekunden). Es musste also weniger ‚Material‘ im Gedächtnis behalten werden. Dies trifft auch für die Anzahl der Noten in der Melodiestimme zu.

Schließlich waren die Teststücke von Dowling mit 20 bis 36 Sekunden deutlich kürzer als in diesem Experiment (48 Sekunden). Es wurde also weniger unterschiedliches Material

präsentiert und die Konzentration der Versuchspersonen für ein Teststück wurde nur halb so lange in Anspruch genommen.

Insgesamt waren die Experimente von Dowling und Mitarbeitern aufgrund der aufgezählten Merkmale wesentlich ‚einfacher‘ für die Versuchsteilnehmer zu bewältigen. Das Ziel jenes Versuches war es vor allem, die Bedeutung von zeitlichen Faktoren (Retentionsintervall und Position des *Testitems* im Stück) auf die Gedächtnisleistung zu evaluieren. Dagegen sollte in diesem Experiment neben zeitlichen Eigenschaften die Bedeutung von einzelnen musikalisch-strukturellen Eigenschaften im Vergessen-/Erinnerungsprozess untersucht werden (die anfangs bezeichneten Modellrahmen zwei und drei). Insofern darf es nicht verwundern, wenn das dreifaktorielle Auswertungsdesign mit Aggregation der Werte einer Vpn. über verschiedene *Testitems* hinweg für die aktuelle Fragestellung keine signifikanten Ergebnisse liefert. Diese Art der Auswertung ist für die im Design angelegte Differenzierung des Testmaterials nach gestaltpsychologischen und melodisch-dimensionalen Merkmalen zu ‚grob‘, d.h. zu ungenau.

Ähnliches gilt auch für das zwei- bzw. dreifaktorielle Modell, das Kauffman und Carlsen (1989) zur Auswertung ihres Gedächtnisexperiments verwendet haben. Zwar aggregieren Kauffman und Carlsen die *ratings* der Versuchspersonen nicht mit Maßen wie der AUC oder d' , aber auch sie stützen sich vor allem auf zeitliche Faktoren und auf die musikalische Vorbildung der Versuchsteilnehmer. Noch reduzierter als Dowling und Kollegen lassen sie die Versuchspersonen nur zwischen identischen und gänzlich verschiedenen *Testitems* unterscheiden. Wenig verwunderlich ist deshalb, dass sie einen *ceiling*-Effekt in ihrem ersten Experiment mit einem Anteil von 84% der korrekten an allen Antworten der Versuchspersonen feststellen. Auch das Design von Kauffman und Carlsen ist deshalb viel zu grob, um strukturelle Unterschiede in den *Testitems* in ihrer Bedeutung für das melodische Gedächtnis untersuchen zu können.

5.5.2 Vierfaktorielles Modell ohne Datenaggregation

Im zweiten Schritt wurde ein Modell gerechnet, welches auf Datenaggregation verzichtet und statt dessen die Rohwerte auf der *Ratingskala* aller Versuchspersonen als abhängige Variable betrachtete. Unabhängige Variablen, waren die drei Hauptfaktoren, unter denen das Experimentaldesign ausbalanciert war (*Position der Referenzmelodie*, *Manipulations-*

klasse, Manipulationstyp) sowie die musikalische Aktivität in drei Stufen. Auch dieses Auswertungsdesign lehnt sich also an das Vorgehen von Dowling und Kollegen an, indem nur Variablen betrachtet werden, die bei der Konstruktion der Teststimuli eine Rolle gespielt haben.

Gegenüber dem Modell mit den aggregierten abhängigen Werten, zeichnet sich dieses Modell dadurch aus, dass der Faktor *Manipulationsklasse* drei statt zwei Stufen besitzt und dass der Faktor *Manipulationstyp* betrachtet werden kann.

Es wurde ein Modell mit allen Haupteffekten sowie allen 2-Wege-Interaktionen gerechnet. Im Sinne der referierten Literatur zu Gedächtnisleistungen für echte Musik, insbesondere der Literatur zu den Akzentmodellen und den Forschungen von Dowling ließen sich folgende Effekte hypothesieren:

Eine Manipulation der Akzente der Testmelodien sollte öfter entdeckt werden als eine Manipulation der nicht-akzentuierten Töne. Dies legen die Akzenttheorien von Boltz und Jones (1986) und Monahan et al. (1987) nahe. Nach den Ergebnissen von Oura und Hatano (1988) ist die Melodiestructur (die hier mit der Akzentuierung gleichgesetzt wird) nur für die musikalisch erfahreneren Versuchspersonen von Bedeutung.

Da viele Studien gezeigt haben, dass das Retentionsintervall bzw. das Material im Retentionsintervall zwischen dem zweimaligen Hören einer Melodie einen Einfluss auf das Wiedererkennen hat (z.B. Kaufmann & Carlsen, 1989; Dowling et al., 2002), sollte die Position der Testmelodie einen Einfluss auf die *ratings* der Versuchspersonen haben.

Nach den Ergebnissen von zahlreiche Forschungen werden die unterschiedlichen Dimensionen einer Melodie unterschiedlich gut erinnert bzw. unterschiedlich schnell vergessen. Auf dem Faktormanipulationstyp sollten demnach Unterschiede zwischen den verschiedenen Faktorstufen festzustellen sein, wobei in der Literatur (z.B. Dowling, 1978; Edworthy, 1985) unterschiedliche Angaben darüber vorliegen, welche Dimensionen besser und welche schlechter erinnert werden.

Die musikalische Aktivität der Versuchspersonen sollte keinen direkten Haupteffekt auf ihre *ratings* ausüben. Allerdings ist nach der Literatur zu erwarten, dass sich Interaktionseffekte z.B. mit der Manipulationsklasse (Bedeutung von Akzenten, s.o.) oder dem Manipulationstyp ergeben.

Das mit den *ratings* der Versuchspersonen gerechnete Modell besaß eine deutlich bessere Modellgüte von $R^2 = 0,199$ und korrigiertem $R^2 = 0,18$ als das obige Modell mit den aggregierten Werten. Wie die folgende Tabelle zeigt, wurden mehrere Haupteffekte und Interaktionen signifikant.

Tests der Zwischensubjekteffekte

Abhängige Variable: complete rating data

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	370,953 ^a	23	16,128	10,502	,000
Konstanter Term	10068,437	1	10068,437	6556,177	,000
MUSICTACT	6,136	2	3,068	1,998	,136
POCOMEL	26,692	1	26,692	17,381	,000
MANICLAS	4,249	1	4,249	2,767	,097
MANTYP	14,885	2	7,443	4,846	,008
MUSICTACT * POCOMEL	15,115	2	7,557	4,921	,007
MUSICTACT * MANICLAS	7,746	2	3,873	2,522	,081
MUSICTACT * MANTYP	7,188	4	1,797	1,170	,322
POCOMEL * MANICLAS	,377	1	,377	,245	,620
POCOMEL * MANTYP	12,149	2	6,074	3,955	,019
MANICLAS * MANTYP	26,551	2	13,275	8,644	,000
Fehler	1491,182	971	1,536		
Gesamt	12980,000	995			
Korrigierte Gesamtvariation	1862,135	994			

a. R-Quadrat = ,199 (korrigiertes R-Quadrat = ,180)

Tabelle 5-8: Lineares Modell mit den *ratings* der Versuchspersonen auf der sechsstufigen Skala als abhängiger Variable. (POCOMEL = Position der Referenzmelodie; MANCOMP = Manipulationsklasse; MUSICTACT = Musikalische Aktivität; MANICLAS = Manipulationsklasse; MANTYP = Manipulationstyp).

Die Richtung der signifikanten Effekte und der teilweise signifikante Einzelvergleiche soll im folgenden stichwortartig wiedergegeben werden. Für die Einzelvergleiche wurde der als konservativ bekannte Scheffé-Test verwendet. Die entsprechenden Tabellen finden sich im Anhang:

Position der Referenzmelodie: Die Melodien im zweiten Teil des Teststückes wurden eher und mit größerer Urteilssicherheit als identisch bewertet als die Melodien aus dem ersten Teil.

Manipulationsklasse: die identischen Melodien hatten nach einem Scheffé-Test im Mittel höhere Identitätswerte auf der *rating*-Skala als die beiden manipulierten Klassen. Die manipulierten Klassen (Akzent und nicht-Akzent) unterschieden sich nicht signifikant.

Manipulationstyp: Neben dem Unterschied zwischen den Melodien ohne Manipulation irgendeines Typs, waren die Werte auf der *rating*-Skala für die im Rhythmus manipulierten

ten Melodien höher als die für die kontur- und intervallmanipulierten Melodien. Die beiden letzteren unterschieden sich nicht signifikant voneinander.

Musikalische Aktivität: Die *rating*-Werte unterschieden sich nicht signifikant nach der musikalischen Aktivität der Versuchspersonen, es zeigten sich aber einige interessante Interaktionen mit diesem Faktor.

*Musikalische Aktivität * Position der Referenzmelodie:* Sowohl die Versuchspersonen mit geringer als auch die mit mittlerer musikalischer Aktivität vergaben höhere Werte für die Melodien aus dem zweiten Teil. Die Versuchspersonen mit hoher musikalischer Aktivität vergaben im Mittel in etwa die gleichen Werte für Testmelodie aus erstem und zweitem Teil.

*Musikalische Aktivität * Manipulationsklasse:* Während sich die Mittelwerte der drei Manipulationsklassen bei den Versuchspersonen mit geringer und mittlerer musikalischer Aktivität nur gering unterschieden (maximale Differenz 0,9), waren sie bei den Versuchspersonen mit hoher musikalischer Aktivität deutlicher verschieden (maximale Differenz 1,6) und befanden sich in der erwarteten Reihenfolge (Akzentmanipulation < Nicht-Akzentmanipulation < keine Manipulation).

*Position Referenzmelodie * Manipulationstyp:* Während die Manipulationen beim Rhythmus und den Intervallen bei den Testmelodien aus dem zweiten Teil des Musikstücks jeweils höhere *rating*-Werte zur Folge hatten, so hatten die Konturmanipulationen bei Melodien aus erstem wie zweitem Teil fast die gleichen *ratings* im Mittel zur Folge.

*Manipulationsklasse * Manipulationstyp:* Bei den akzentmanipulierten Melodien waren die veränderten Intervalle das deutlichste Anzeichen für eine veränderte Melodie, d.h. sie bekamen im Mittel die niedrigsten *rating*-Werte. Bei den nicht-akzentmanipulierten war es dagegen die veränderte Kontur.

Das Modell bestätigt also weitgehend die vorab hypothetisierten Annahmen und damit die entsprechenden Ergebnisse in der Literatur (vgl. auch die Ergebnistabellen der Unterkapitel 3.2 und 3.3).

Obwohl dieses Modell noch weit davon entfernt ist, die *ratings* der Versuchspersonen vollständig zu erklären (korrigiertes $R^2 = 0,18$), so lassen sich anhand der gerade beschriebenen signifikanten Effekte doch einige interessante Erkenntnisse gewinnen:

Für die musikalisch weniger Aktiven spielt die Position der Referenzmelodie eine sehr wichtige Rolle: Ist wenig Zeit vergangen, seitdem die Testmelodie im Stück erklingen ist, so sind sie eher geneigt, sie als identisch zu bewerten. Die möglichen Manipulationen werden dabei eher ‚übersehen‘ als bei den musikalisch Aktiveren. Dies gilt jedoch nicht für die Kontur. Eine Interpretation dafür wäre, dass die musikalisch Aktiveren ein besseres Gedächtnis für Detailunterschiede in den Melodien haben, während die andere Gruppe Unterschiede mit Ausnahme der Melodiekontur nicht wahrnimmt bzw. schnell vergisst. Die größere Bedeutung der Melodiekontur für Personen mit geringerer bzw. mittlerer musikalischer Vorbildung hatte auch schon eine Vielzahl von früheren Studien finden können (z.B. Dowling & Hollombe, 1977; Dowling, 1986).

Dazu passend ergaben sich die erwarteten Unterschiede zwischen akzentmanipulierten und nicht akzent-manipulierten Melodien nur für die Gruppe der musikalisch Aktiven. Für die weniger Aktiven spielte es keine Rolle, welche Töne einer Melodie manipuliert wurden (im Sinne der verwendeten heuristischen Akzentdefinition). Dies passt zu den Ergebnissen von Oura und Hatano (1988) oder Dowling et al. (1995), die für musikalisch vorgebildete Versuchspersonen eine Orientierung an der Melodiestructur im Melodiegedächtnis gefunden hatten, während die nicht-vorgebildeten Versuchspersonen nur schwer zu systematisierende Gedächtnisfehler zeigten.

Unabhängig von der musikalischen Aktivität wird der exakte Rhythmus der Melodien am schnellsten vergessen.

Eine interessante Interaktion zwischen *Manipulationsklasse* und *-typ* zeigt, dass die Intervalle vor allem dann im Gedächtnis bleiben, wenn mindestens ein akzentuierter Ton an dem Intervall beteiligt ist. Die musikalische Kontur wird dagegen auch auf nicht-akzentuierten Tönen beachtet und prägt sich auch dann ins Gedächtnis ein.

Der Wechsel der Modellperspektive von einem Modell, welches die Gedächtnisleistungen in Form von aggregierten Daten erklärt, zu einem welches die *ratings* der Versuchspersonen als Rohdaten betrachtet, bietet einige Vorteile:

Beim erstgenannten Modell waren keinerlei signifikante Haupteffekte und Wechselwirkungen zu beobachten, woraus sich folgern lässt, dass die betrachteten Faktoren keinen nachweisbaren Einfluss auf die Gedächtnisleistungen hatten. Ein Grund für die fehlenden Effekte ist mit Sicherheit die Nicht-Berücksichtigung des Manipulationstyps als Faktor

aufgrund der hier nicht durchführbaren Datenaggregation. Eine differenziertere Betrachtung der entsprechenden Wechselwirkungen dieses Faktors ist damit nicht möglich.

Das zweite Modell zeigt, dass sich die Antworten der Versuchspersonen nicht zufällig verteilen, sondern dass alle betrachteten Faktoren eine Rolle für die *ratings* spielen. Beispielsweise werden zwar die *ratings* nicht ‚richtiger‘, d.h. die Gedächtnisleistungen nicht besser, wenn die Testmelodie aus dem zweiten Teil des Stückes stammt. Die *ratings* werden insgesamt aber signifikant höher. Auch hieraus lässt sich eine Erkenntnis über das menschliche Melodiegedächtnis ziehen: Je mehr Zeit vergangen ist bzw. je mehr andere Melodien man in der Zwischenzeit hört, desto eher wird eine früher gehörte Melodie vergessen, und gleichzeitig reicht das Wiedererkennen einer nur teilweise identischen Melodie oft schon aus, um eine Melodie als identisch zu werten. Dies gilt zumindest für die musikalisch weniger Aktiven.

Anstatt jedoch das vierfaktorielle Modell bis ins letzte Detail zu analysieren und zu interpretieren, soll der Ansatz der Erklärung der *rating*-Daten der Versuchspersonen noch ein Stück weitergegangen werden:

Der immer noch geringe R^2 -Wert des vierfaktoriellen Modells legt nahe, dass hierin noch einige Faktoren unberücksichtigt geblieben sind, die die *ratings* der Versuchspersonen beeinflussen. Da es sich bei dem verwendeten Testmaterial um echte Popstücke mit ‚natürlichen‘ Melodien und einer Begleitung handelte, ist zu vermuten, dass die nicht im Design kontrollierbaren Eigenschaften der Melodien selbst eine gewichtige Rolle in den kognitiven Prozessen gespielt haben, die bei der Teilnahme an dem Versuch notwendig waren. Außerdem soll ein Modell mit Prädiktoren gefunden werden, welche sich nicht aus dem Vorwissen des Experimentators ableiten. Dies betrifft z.B. Manipulationsklasse und Manipulationstyp. Können diese Prädiktoren durch solche ersetzt werden, die die Relation hinsichtlich der Manipulation zwischen zwei Melodien auch in nicht kontrollierten Situationen wiedergeben, so ist die Generalisierbarkeit des Modells für einen größeren Bereich von Situationen gegeben. Dies sollte idealerweise typische *Query-by-humming*-Situationen einschließen.

Die nachfolgende statistische Modellierung versucht, diesen bisher nicht berücksichtigten Eigenschaften des Testmaterials Rechnung zu tragen.

5.6 Statistische Modellierung der *Versuchspersonenratings* (*data-mining-Ansätze*)

Der Ansatz der folgenden statistischen Modellierungen folgt dem typischen *data-mining*-Prinzip, aus einem Maximum an verfügbaren Prädiktoren diejenigen zu selektieren und zu kombinieren, die die Reaktionen der Versuchspersonen auf der *rating*-Skala am besten erklären. Dabei werden im Prinzip die oben erläuterten 13 Variablen aus den drei verschiedenen Gruppen bzw. Modellbereichen herangezogen sowie einige Anpassungen an den vier bereits im vorigen Modell verwendeten Variablen vorgenommen (s.u.).

Die Variablenselektion und –kombination in einem Modell soll mit zwei verschiedenen Ansätzen geleistet werden. Der erste Ansatz sucht nach einem allgemeinen linearen (Regression-)Modell der bekannten Form

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + e_i,$$

welches die *ratings* der Ver-

suchspersonen als abhängige Variable y vorherzusagen sucht. Der zweite Ansatz bedient sich der aus dem Bereich des Maschinlernens bekannten Klassifizierungs- und Regressionsbaummodelle, um die *ratings* der Versuchspersonen zu prognostizieren.

Um die Modelle in ihrer Voraussagekraft anschließend vergleichen zu können, wird der Datensatz der insgesamt 995 Beobachtungen in mehrere Teile geteilt. Für das lineare Modell wird die Variablenselektion auf einem Drittel der Beobachtungen vorgenommen. Ein weiteres Drittel der Daten steht für die Schätzung der Modellparameter bereit. Auf dem letzten Drittel soll schließlich die Vorhersageleistung des Modells getestet werden. Zur Entwicklung der Baummodelle stehen zwei Drittel der Beobachtungen zur Verfügung. Die Vorhersagequalitäten der Baummodelle sollen schließlich auf dem selben Datensatz geprüft werden, der auch für die Testung des linearen Modells zur Verfügung steht. Der gesamte Datensatz wurde anhand von Zufallsziehungen von je 14 aus 42 Versuchspersonen vorgenommen, d.h. alle Beobachtungen (*ratings*) einer Vpn. waren in einem Drittel des Datensatzes enthalten.

Terminologisch ist zu beachten, dass der Begriff *data mining*, hier nicht für die Beschreibung der statistischen Modelle verwendet wird. Denn in der statistischen Literatur steht *data mining* in der Regel für neue oder gerade aktuelle Ansätze der Datenanalyse. Lineare Modelle und auch Baummodelle sind jedoch schon seit einigen Jahrzehnten bekannt. Mit *data mining* ist im Zusammenhang dieser Arbeit dagegen vielmehr der explorative Ansatz gemeint, aus einer Vielzahl von messbaren Größen ein Erklärungsmodell für eine kogni-

tive musikbezogene Leistung zu konstruieren. Dieser Ansatz steht im Gegensatz zu der in der Musikpsychologie vorherrschenden Hypothesenprüfung, bei der erwartete Effekte einer oder weniger Einflussgrößen auf eine musikalische Leistung getestet werden. Der gerade mit dem Begriff *data mining* beschriebene Ansatz scheint für die Analyse dieses Experimentes jedoch aussichtreicher, da durch die angestrebte größere ökologische Validität der Versuchsbedingungen eine Vielzahl von Eigenschaften der ‚echten‘ Musikstücke und der quasi realen Hörsituation Einflüsse auf die Gedächtnisleistungen der Probanden haben mögen.

5.6.1 Variablenkodierung und Operationalisierung

Die Anpassung der im obigen Modell verwendeten Variablen und Einflussgrößen schien vor der eigentlichen statistischen Modellierung aus zwei Gründen geboten. Zum einen sind Prädiktoren, die alle ein vergleichbares Skalenniveau besitzen, in den meisten Softwarepaketen einfacher einzusetzen als eine Auswahl von Prädiktoren, die gleichzeitig mehrstufige nominale Variablen und solche mit ordinalem und metrischen Skalenniveau vereinigt. Es sollten also die Einflussgrößen so operationalisiert werden, dass sie entweder metrisches Skalenniveau besitzen oder dichotom ausgeprägt sind, so dass sie in jedem Fall als Prädiktor in eine lineare Regression aufgenommen werden können. Das betrifft die *Manipulationsklasse*, den *Manipulationstyp* und die dreistufigen ordinalen Variablen zum musikalischen Hintergrund.

Zum anderen sollten die Variablen so operationalisiert werden, dass sie auch für zukünftige Anwendungen außerhalb eines kontrollierten experimentellen Designs eingesetzt werden können. Denn nicht immer ist im vorhinein klar, ob die Kontur oder der Rhythmus einer Melodievariante von dem Original abweicht oder ob die Akzente der Melodie betroffen sind.

Bezüglich der beiden Variablen zum musikalischen Hintergrund wurden die metrischen statt der ordinalen Varianten verwendet.

Die kategoriale Variable *Manipulationstyp* wurde in drei dichotome Dummyvariablen kodiert, die jeweils angeben, ob eine Manipulation in einer der drei melodischen Dimensionen Rhythmus, Interval, Kontur, vorliegt oder nicht.

Die Aufbereitung und Zusammenfassung von *Manipulationsklasse* und *Akzentgewicht* zu einer einzigen metrischen Variablen gestaltet sich dagegen etwas aufwendiger. Das Vorgehen soll im folgenden Exkurs skizziert werden.

Die vollständige Liste der als Prädiktoren verwendbaren Variablen findet sich am Ende des Exkurses zur Akzentähnlichkeit.

5.6.1.1 Exkurs: Konstruktion des Prädiktors *Akzentähnlichkeit*

Die Variablen *Akzentgewicht* und *Manipulationsklasse* wurde in diesem Versuch durch den Versuchsleiter festgelegt bzw. das musikalische Testmaterial wurde nach Maßgabe einer bestimmten Verteilung dieser Variablen konstruiert. Diese beiden Variablen sollen durch eine einzige Größe *Akzentähnlichkeit* ersetzt werden, die durch Messung an dem melodischen Material erhoben werden kann. Dieser Ansatz, ein Messinstrument für die Ähnlichkeit der Akzentstrukturen zu konstruieren, trägt der Tatsache Rechnung, dass in einer beliebigen Situation, in der die Gedächtnisleistung für eine Melodie vorhergesagt werden soll, nicht im vorhinein klar ist, in welchem Maß melodische Akzente von Veränderungen betroffen sind und welche melodischen Dimensionen im Gedächtnis verändert werden. Die Messung der Akzentähnlichkeit soll mittels eines algorithmischen Verfahrens an beliebigen einstimmigen Melodien möglich sein. Ausgangspunkt für die Konstruktion dieser Variable sind zum einen die in Kapitel 3.3 vorgestellten Akzenttheorien und –regeln von Monahan, Boltz und Jones und anderen sowie das Optimierungsverfahren zu algorithmischen Ähnlichkeitsmaßen für Melodien, wie es in Kapitel 4 beschrieben ist.

Zwar ist auch bei Large et al. (1995) ein offenbar brauchbarer Mechanismus zur Reduktion von Melodien auf eine Skelett- oder Akzentstruktur in Hinsicht auf Gedächtnisprozesse beschrieben worden, jedoch basiert dieser auf einem vollkommen anderen komputationellen Ansatz, nämlich einer bestimmten Klasse von neuronalen Netzen. Das verwendete Netz muss in jedem Fall zuvor mit Melodien eines bestimmten Repertoires trainiert werden und seine universelle Einsetzbarkeit ist bisher weder nachgewiesen noch behauptet worden. Die im folgenden verwendeten Akzentregeln basieren dagegen auf Gestaltgesetzen, die für die menschliche Kognition an sich Geltung beanspruchen. Zudem sind sie weit besser in den hier verwendeten regelbasierten, algorithmischen Ansatz einzubetten als die neuronale Architektur, die Large und Kollegen vorgeschlagen haben.

Für die Konstruktion der manipulierten Testmelodien in diesem Experiment war eine bestimmte Operationalisierung der Akzentregeln von Monahan und Boltz und Jones gewählt worden, wie sie im Abschnitt Methode unter 5.1.3 beschrieben ist. Dort wurde bereits darauf hingewiesen, dass die gewählte Operationalisierung zwar plausibel erschien, aber kei-

neswegs die einzig mögliche Wahl darstellt und keine Garantie existiert, dass durch diese Operationalisierung die ‚wahre‘ Akzentstruktur der Testmelodien beschrieben wird, wie sie die Versuchspersonen wahrnehmen. Diese Unsicherheit ist durch die Uneindeutigkeit der Regelformulierungen in der Literatur begründet. Zum Beispiel ist in der folgenden Formulierung nicht eindeutig definiert, was eine lange oder kurze Dauer (im Vergleich zum umgebenden Kontext) ist: „Temporal accents occur on tones following a lengthened (or shortened) duration such as a pause.“ (Boltz & Jones, 1986, S. 407). Ist eine lange Dauer eine, die länger als die Vorgängerdauer ist oder zweimal so lang wie die Vorgängerdauer oder länger als der Modus der Tondauern (z.B. Achtel) einer Melodie? In der Literatur finden sich dazu keine definitiven Aussagen. Vorteil dieser Unbestimmtheit ist, dass diese Akzentregeln nicht wirklich empirisch falsifiziert werden können.

Um die optimale Operationalisierung der Akzentregeln zu finden wurde der simple empirische Ansatz gewählt, der sich auch schon bei der Optimierung der Ähnlichkeitsmaße als tragfähig erwiesen hatte: Es wird eine Vielzahl von möglichen Regeln und Regelkombinationen gegen empirisch erhobene Daten getestet. Die Regelkombination(en) mit der besten Vorhersagekraft werden als Annäherung an die Beschreibung der ‚wahren‘ Akzentstruktur genommen.

Dieser Ansatz geht von mehreren Prämissen aus:

- Je ähnlicher sich die Akzentstrukturen von zwei Melodien sind, desto höher wird die Gesamtähnlichkeit der Melodien von Experten eingeschätzt. Dafür, dass die Akzentstruktur von Melodien v.a. für musikalische Vorgebildete bzw. Experten relevant sind, lassen sich zahlreiche Hinweise in der Literatur finden (z.B. Monahan et al., 1987; Oura & Hatano, 1988; Smith & Cuddy, 1989). Die *Ähnlichkeitsratings* der Experten aus dem ersten Experiment (Kapitel 4) können deshalb als empirische Daten verwendet werden. Das Auffinden der ‚wahren‘ Akzentstruktur wird als Optimierung von Ähnlichkeitsmaßen formuliert, die auf Akzentregeln beruhen.
- Nur die Kombination von Akzenten aus mehreren musikalischen Dimensionen (harmonisch, melodisch, rhythmisch etc.) kann eine adäquate Repräsentation der globalen Akzentstruktur einer Melodie erzeugen (nicht umsonst heißt die Hypothese von Boltz & Jones „joint accent hypothesis“). Aus pragmatischen Gründen wurden hier Akzentkombination aus vier, fünf, sechs und sieben Dimensionen betrachtet.

- Aus jeder musikalischen Dimension kann nur eine Regel in die Kombination mit Regeln aus anderen Dimensionen eingehen. Die Regeln innerhalb einer Klasse sind alternativ, die aus verschiedenen Klassen additiv zu behandeln (Ausnahme s.u.).
- Die Addition von Akzentwerten ist als Verfahren für die Konstruktion einer globalen Akzentstruktur angemessen.

Nimmt man diese Prämissen an, so ergeben sich als Vorgehen zum Auffinden der optimalen Kombination von Akzentregeln folgende Schritte:

- A. Exakte Definition aller möglichen Akzentregeln und Varianten ausgehend von den in der Literatur unvollständig beschriebenen Regeln.
- B. Ermittlung der Akzentstrukturen aller Melodien und Varianten der Experimente 1 und 3 aus Kapitel 4 für jede der 25 Akzentregeln einzeln und Berechnung der Akzentähnlichkeiten für jedes Melodie-Variante-Paar für jede Akzentregel mit der Edit Distance.
- C. Berechnung der Distanzen zwischen den Akzentähnlichkeiten jeder Regel und den *Ähnlichkeitsratings* der Versuchspersonen Ermittlung der ‚besten‘ Regeln mit der geringsten Distanz.
- D. Kombination der ‚besten‘ Regeln zu plausiblen *joint accent measures*.
- E. Berechnung der Ähnlichkeit zwischen den gemeinsamen Akzentstrukturen der Melodien und Varianten nach mehreren verschiedenen Berechnungsverfahren. Berechnung der Distanz zwischen Ähnlichkeitswerten der *joint accents measures* und den mittleren *ratings* der Experten.
- F. Ergebnisse der Distanzberechnung und Auswahl der besten Maße als Prädiktoren für das Modell zur Erklärung der *ratings* in diesem Wiedererkennensexperiment.

A. Exakte Definition der verwendeten Akzentregeln:

Zunächst müssen die Akzentregeln, die miteinander kombiniert werden sollen, exakt definiert und einer musikalischen Dimension eindeutig zugeordnet werden. Als Ausgangsbasis dienen die von Monahan et al. (1987) und Boltz & Jones (1986) aufgestellten Regeln, die selbst auf frühere Forschungen rekurren (z.B. Yeston, 1976; Thomassen, 1982; Povel & Essens, 1985). Zu jeder in der Literatur genannten Regel sind sinnvolle und plausible Varianten in der Operationalisierung angegeben. Die Regeln selbst sind mit einem höchstens achtstelligen Kurznamen versehen und fett gedruckt. Bei den Regeln, die nur bei einem

Autor in der Literatur auftauchen, ist die Quelle der ursprünglichen Formulierung in Klammern vermerkt. Die 25 formulierten Regeln wurden insgesamt sieben verschiedenen Klassen zugeordnet, die die unterschiedlichen musikalischen Dimensionen repräsentieren, die bei der Betrachtung von Melodieakzenten eine Rolle spielen können.

Voraussetzung für die Anwendung der Regeln ist in vielen Fällen die Einteilung der Melodien in einzelne Phrasen. Diese Phraseneinteilung wurde (wie oben beschrieben) anhand langer *Interonset*-Intervalle vorgenommen: Eine neue Phrase beginnt auf einem Ton, wenn die Dauer des Vorgängertons mindestens viermal so lang wie der Modus der Tondauern der Melodie ist.

Klasse I Sprung:

jumpaft3, jumpaft4, jumpaft5 (jump after): Jeder Ton, dem ein Intervall von $\geq 3(4,5)$ Halbtönen vorausgeht, bekommt einen Akzent.

jumpbea3, jumpbea4, jumpbea5 (jump before and after): Jeder Ton, dem Intervall von $\geq 3(4,5)$ Halbtönen vorausgeht oder folgt, bekommt einen Akzent.

jumploc (jump local): Jeder Ton, dem ein Intervall vorausgeht, das um mindestens 2 Halbtöne größer ist, als sein Vorgänger- und sein Nachfolgerintervall, bekommt einen Akzent (geht auf zurück auf Monahan et al., 1987).

shortphr (short phrase): Wenn eine Phrase aus nur 2 Tönen besteht, gib dem 2. Ton einen Akzent (nur bei Phraseneinteilung).

Klasse II Konturumkehr:

pextrem (pitch extremum): Gib jedem Ton, dessen Vorgänger- und Nachfolgeton jeweils höher oder tiefer ist, einen Akzent.

pextrst (pitch extremum Steinbeck): Gib jedem Ton, dessen Vorgänger- und Nachfolgeton jeweils höher oder tiefer ist und der keine Wechselnote (sensu Steinbeck, s. Definition in Kapitel 4) ist, einen Akzent.

pextrmf (pitch extremum Müllensiefen&Frieler): Gib jedem Ton, dessen Vorgänger- und Nachfolgeton jeweils höher oder tiefer ist und der keine Weselnote (sensu Müllensiefen & Frieler, s. Definition in Kapitel 4) ist, einen Akzent.

pextrsta (after pitch extremum Steinbeck): Gib jedem Ton, der auf eine Konturumkehr (sensu Steinbeck) folgt, einen Akzent (geht auf zurück auf Boltz & Jones, 1986).

Klasse III längere Dauer:

longpr (longer previous): Gib jedem Ton, dessen Dauer länger ist als die seines Vorgängertons, einen Akzent.

long2pr (2x longer previous): Gib jedem Ton mit Dauer $\geq 2x$ Vorgängerton einen Akzent.

longmod (longer modus): Gib jedem Ton, dessen Dauer länger ist als der Modus der Dauern, einen Akzent.

long2mod (longer 2x modus): Gib jedem Ton, dessen Dauer länger ist als 2x der Modus der Dauern, einen Akzent.

Klasse IV, kürzere Dauer:

shortpr (shorter previous): Gib jedem Ton mit Dauer $< \text{IOI}$ des Vorgängertons einen rhythmischen Akzent, wenn dieser Ton der Anfang einer Tongruppe von drei oder mehr Tönen ist, d.h. keine Phrasengrenze erreicht ist; nur bei Phraseneinteilung. (Monahan et al., 1987)

short2pr (2xshorter previous): Gib jedem Ton mit Dauer $\leq 0,5x$ Dauer des Vorgängertons einen rhythmischen Akzent, wenn dieser Ton der Anfang einer Tongruppe von drei oder mehr Tönen ist, d.h. keine Phrasengrenze erreicht ist; nur bei Phraseneinteilung. (Monahan et al., 1987)

endloioi (end long IOI): Gib jedem Ton, dem ein IOI mit $\geq 2x$ Modus der Tondauern vorausgeht und dessen Vorgängerton keinen Akzent nach **longpr** bekommen hat, einen Akzent. (Boltz & Jones, 1986)

Klasse V Metrum:

beats13 (beats 1 + 3): Gib jedem Ton, der sein Onset auf den betonten Taktzählzeiten (1 + 3) hat, einen metrischen Akzent.

beat1 (beat 1): Gib jedem Ton, der sein Onset auf der Takteins hat, einen metrischen Akzent.

Klasse VI Harmonik:

triad (notes of triad of bar's main tonality): Gib jedem Ton, der mit Grundton, Terz oder Quinte des im Takt zugrundeliegenden Akkordes identisch ist, einen Akzent. (Tonalitätsbestimmung wie beim Ähnlichkeitsmaß *harmcore*)

triaphen (notes on phrase ends of triad of bar's main tonality): Gib jedem Ton auf einem Phrasen- oder dem Melodieende, der mit Grundton, Terz oder Quinte des im Takt zugrundeliegenden Akkordes identisch ist, einen Akzent (Boltz, 1999)

Klasse VII Ränder (Regeln können auch additiv verwendet werden):

phrasbeg (phrase beginning): Gib jedem Ton, der eine Phrase beginnt, einen Akzent.

phrasend (phrase ending): Gib jedem Ton, der eine Phrase beendet, einen Akzent.

Natürlich stellen auch diese 25 Regeln keine vollständige Sammlung aller möglichen Akzentregeln für Melodien dar. Dennoch decken sie den überwiegenden Teil der in der Literatur genannten Regeln ab.

B. Berechnung der Akzentähnlichkeit anhand der Einzelregeln:

Da die kombinatorischen Möglichkeiten zur Verbindung der Akzentregeln in *joint accent measures* enorm sind (die Anzahl von Maßen aus genau sieben Einzelregeln ist schon 6144, wenn die Regeln für die Ränder nur alternativ verwendet werden), bedarf es eines Selektionsmechanismus, der die zu testenden kombinierten Maße auf eine überschaubare Anzahl reduziert.

Da für die vorliegende Aufgabenstellung innerhalb des Wiedererkennensexperimentes der Vergleich von einem Melodieoriginal mit einer zugehörigen Variante von Interesse ist, wurden nur die Daten des ersten und des dritten Ähnlichkeitsexperimentes aus Kapitel 4 verwendet innerhalb dieses Selektions- und Optimierungsprozesses verwendet. Es wurden für jede Melodie der 84 bzw. 32 Melodiepaare die Akzentstruktur nach jeder der 25 Akzentregeln bestimmt und die Ähnlichkeit für jedes Paar mit Hilfe der Edit Distance berechnet. Die Edit Distance wurde an dieser Stelle als Verfahren für die Berechnung der Ähnlichkeit gewählt, da sie sich bei der globalen Betrachtung der Ähnlichkeitsalgorithmen in Kapitel 4 als universellstes und leistungsstärkstes Verfahren erwiesen hatte. Da die Akzentstrukturen nach Anwendung jeder Einzelregel nur aus einer Serie von Nullen und Einsen bestehen (0 = kein Akzent; 1 = Akzent), ist die einfache Edit Distance ein geeignetes

Maß, dass sie auch beider Verwendung von Symbolalphabeten mit nur wenigen Symbolen sinnvolle Ergebnisse liefern kann und keine Strings oder Vektoren gleicher Länge für den Vergleich voraussetzt.

Zur Illustration soll hier die Melodie des Titels „Wonderland“ von Passion Fruit dienen, die auch schon in Kapitel 4.1.3 als Beispiel verwendet wurde und dort in Notenform dargestellt ist. Nach Anwendung der Regel *long2pr* ergibt sich folgende Serie von Nicht-Akzenten und Akzenten:

„Wonderland“ (Original) transformiert durch *long2pr*:
000010100010000101000100001010001000001000001001

C. Berechnung der Distanzen zwischen Akzentähnlichkeiten der Einzelregeln und den *ratings* der Versuchspersonen:

Genauso wie an analoger Stelle in Kapitel 4 wurde nun die Übereinstimmung zwischen den Mittelwerten der *Ähnlichkeitsratings* der Versuchsteilnehmer und den 25 algorithmischen Akzentregeln mit Hilfe der euklidischen Distanz über die z-transformierten Ähnlichkeitswerte berechnet.

Die Ergebnisse für die Daten des ersten und des dritten Experimentes unterschieden sich kaum. Die jeweils hochsignifikanten Korrelationen zwischen den Distanzreihen der 25 Akzentregeln waren 0,847 (Pearson) und 0,815 (Spearman).

Die besten Einzelmaße aus den sieben Akzentklassen waren für die Daten von Experiment 1:

jumpaft4, *pextrmf*, *long2pr*, *endloioi*, *beat13*, *triad*, *phrasbeg* / *phrasend* (*phrasebeg* und *phrasend* erreichten den selben Distanzwert).

Für die Daten von Experiment 3:

jumpaft4, *pextrst*, *long2mod*, *enloioi*, *beat1*, *triad*, *phrasbeg* / *phrasend* (s.o.).

Da die Distanzen zu den *ratings* aus Experiment 1 auf dem größeren Datensatz beruhen und das Urteil von mehr Experten wiedergeben, wurden im Falle der zwei abweichenden besten Maße in einer Gruppe (Konturumkehr und Metrum) die besten Maße gemäß Experiment 1 gewählt. Zusätzlich wurde als Alternative zu *jumpaft4*, das in beiden Distanztabelle nur wenig schlechtere *jumpbea3* mit in die Auswahl der besten Maße aufgenommen, da sich in einer Untersuchung von Frieler (2005) eine ähnliche Regel als effizient erwiesen hatte. Da die beiden Regeln *shortphr* und *phrasbeg* vor allem dann überhaupt

angewendet werden können, wenn eine Melodie viele kurze Phrasen enthält, sind sie als gemeinsames Regelpaar definiert und getestet worden.

Die für die Kombination innerhalb der *joint accent measures* ausgewählten Maße sind also: *jumpaft4*, *jumpbea3*, *pextrmf*, *long2pr*, *endloioi*, *beat13*, *triad*, *phrasend*, *shortphr+phrasbeg*.

D. Konstruktion der joint accent measures durch Kombination:

Auch die Kombination der besten Maße zu *joint accent measures* wurde durch Heuristiken geleitet, um der kombinatorischen Möglichkeiten Herr zu werden. Diese Heuristiken waren:

- Ein *joint accent measure* sollte mindestens vier Maßen bestehen, die unterschiedliche Klassen repräsentieren, denn es wird angenommen, dass die gemeinsame Akzentstruktur für die menschliche Wahrnehmung relevant ist.
- Ein Maß aus den Tonhöhenklassen Sprung oder Konturumkehr muss in jedem Maß enthalten sein, denn die Tonhöhendimension wird sowohl in der Literatur als wichtigste Dimension bezeichnet als auch hatte sie sich in den Ähnlichkeitsexperimenten in Kapitel 4 als bedeutendste Dimension erwiesen.
- Ein Dauernmaß aus den Klassen längere Dauer oder kürzere Dauer muss in jedem *joint*-Maß enthalten sein, denn alle Theorien zu gemeinsamen Akzenten gehen von der bedeutsamen Interaktion zwischen Tonhöhen- und Dauernakzenten aus.
- Bei der Konstruktion von ‚4er‘-Maßen sollen die Maße einbezogen werden, die inhaltlich vermutlich am deutlichsten verschieden sind. D.h. zu den beiden bereits gesetzten Maßen aus einer Tonhöhen- und einer Dauerndimension kommen Metrum und Harmonik hinzu (und nicht die Maße aus der jeweils anderen Tonhöhen- und Dauerndimension).
- Da über die Bedeutung des Regelpaars *shortphr+phrasbeg* im Kontext der übrigen Regeln keine begründbaren Vermutungen gemacht werden können, werden alle Maße einmal mit und einmal ohne dieses Regelpaar getestet.

Die Liste der mit Hilfe dieser Heuristiken generierten 18 Maße ist die folgende. Das Regelpaar *shortphr+phrasbeg* ist hier aus Platzgründen nicht aufgelistet. Alle Maße dieser Liste sind jedoch noch einmal mit diesem Regelpaar zusammengesetzt worden, was die Liste verdoppelt (insgesamt 36 Maße).

Joint accent measure	Beteiligte Regeln					
JOINT41	JUMPAFT4	LONG2PR	BEAT13	TRIAD		
JOINT42	PEXTRMF	LONG2PR	BEAT13	TRIAD		
JOINT43	JUMPAFT4	ENDLOIOI	BEAT13	TRIAD		
JOINT44	PEXTRMF	ENDLOIOI	BEAT13	TRIAD		
JOINT45	JUMPBEA3	LONG2PR	BEAT13	TRIAD		
JOINT46	JUMPBEA3	ENDLOIOI	BEAT13	TRIAD		
JOINT51	JUMPAFT4	PEXTRMF	LONG2PR	BEAT13	TRIAD	
JOINT52	JUMPAFT4	PEXTRMF	LONG2PR	PHRASEND	TRIAD	
JOINT53	JUMPAFT4	PEXTRMF	ENDLOIOI	BEAT13	TRIAD	
JOINT54	JUMPAFT4	PEXTRMF	ENDLOIOI	PHRASEND	TRIAD	
JOINT55	JUMPBEA3	PEXTRMF	LONG2PR	BEAT13	TRIAD	
JOINT56	JUMPBEA3	PEXTRMF	LONG2PR	PHRASEND	TRIAD	
JOINT57	JUMPBEA3	PEXTRMF	ENDLOIOI	BEAT13	TRIAD	
JOINT58	JUMPBEA3	PEXTRMF	ENDLOIOI	PHRASEND	TRIAD	
JOINT61	JUMPAFT4	PEXTRMF	LONG2PR	BEAT13	PHRASEND	TRIAD
JOINT62	JUMPAFT4	PEXTRMF	ENDLOIOI	BEAT13	PHRASEND	TRIAD
JOINT63	JUMPBEA3	PEXTRMF	LONG2PR	BEAT13	PHRASEND	TRIAD
JOINT64	JUMPBEA3	PEXTRMF	ENDLOIOI	BEAT13	PHRASEND	TRIAD

Tabelle 5-9: Liste der getesteten *joint accent measures* und der beteiligten Akzentregeln

Bei der Liste der *joint accent measures* ist zu bedenken, dass dies durch die Einschränkungen der Heuristiken natürlich keine vollständige, aber eine plausible Liste von Regelkombinationen ist. Es ist natürlich möglich, dass die optimale Kombination von Akzentregeln sich nicht in dieser Liste findet. Jedoch ist es wahrscheinlicher aus einer Menge von 36 plausiblen Regelkombinationen eine optimale oder zumindest effiziente Kombination wirklich zu erhalten als wenn – wie in der Literatur üblich – nur eine Regelkombination als effizienteste postuliert wird.

Die Akzentwerte aus Anwendung der verschiedenen Regeln werden im *joint accent measure* addiert. Natürlich ist auch eine Gewichtung der einzelnen Regeln denkbar, doch es liegen in der Literatur nur sehr wenige Hinweise zu einer Gewichtung vor, und diese wenigen Hinweise enthalten dabei keine konkreten Gewichtsverhältnisse. Aus diesem Grund werden die Akzente der verschiedenen Regel zunächst alle mit 1 gewichtet. Zur Illustration der Akzentstruktur an diesem Punkt nach Transformation durch die *joint*-Maße soll wieder die Melodie von „Wonderland“ dienen.

„Wonderland“ (Original) transformiert durch *joint41*:

101142201141011422001410113120013121013121013013

Es lassen sich also schon deutlich Unterschiede zwischen den Akzentgewichten der einzelnen Töne ausmachen.

E. Ermittlung der Akzentstrukturen durch die *joint accent measures* und Vergleich mit den *Expertenratings*:

Wie unter Punkt B) und C) für die Einzelregeln beschrieben wurden nun zunächst die Ähnlichkeiten zwischen den Melodie-Varianten-Paaren der Experimente 1 und 3 aus Kapitel 4 mit jedem *joint accent measure* berechnet und anschließend die Distanzen dieser Ähnlichkeitswerte zu den mittleren Urteilen der Experten berechnet.

Im Gegensatz zur Berechnung der Ähnlichkeit zwischen zwei Akzentreihen, die durch die Anwendung von nur einer Regel entstanden, ist beim der Vergleich zweier Akzentreihen, die durch die Transformation der *joint measures* zustande gekommen sind, die einfache Edit Distance nicht mehr das einzig mögliche bzw. plausible Vergleichsverfahren.

Da die Symbole der Akzentreihe als Elemente einer gerichtete Größe im Sinne der Dimension Akzentgewicht interpretiert werden können, scheint nun auch eine Betrachtung der Unterschiede zwischen den Symbolen bzw. Symbolreihen sinnvoll. Die Unterschiede können durch eine Klassifizierung der Akzentgewichte in Akzentklassen oder durch eine Gewichtung der Operationen der Edit Distance anhand der Akzentgewichte berücksichtigt werden. Für den Vergleich zweier Akzentreihen wurden sechs verschiedene Varianten der Edit Distance eingesetzt:

0. einfache Edit Distance 1: jedes Akzentgewicht ist eigenes Symbol ($n+1$ Symbole bei n verschiedenen Akzentgewichten); Kosten für Löschen und Ersetzen: 1
1. einfache Edit Distance 2: Einteilung in Akzente und Nicht-Akzente (2 Symbole); Kosten für Löschen und Ersetzen: 1
2. einfache Edit Distance 3: Einteilung in starke Akzente (Akzentgewicht 3+) und nicht-Akzente (Akzentgewicht 0), 2 Symbole; Kosten für Löschen und Ersetzen: 1;
3. einfache Edit Distance 4: Einteilung in schwache (Akzentgewichte: 1, 2), starke Akzente (Akzentgewichte: 3+) und nicht-Akzente (Akzentgewicht: 0), Kosten für Löschen und Ersetzen: 1

4. gewichtete Edit Distance 1: Differenz zwischen Akzentgewichten als Kosten, kein Akzent = 0
5. gewichtete Edit Distance 2: Differenz zwischen Akzentklassen (0, 1, 2; s. Einteilung bei 3.) sind Kosten

Da jede der 36 oben aufgelisteten Regelkombinationen mit jeder der sechs Edit Distance Varianten kombiniert wurde, ergaben sich schließlich 216 *joint accent measures*, die empirisch mit den *Expertenratings* verglichen wurden.

F. Ergebnisse der Distanzberechnungen und Auswahl des besten Maßes als Prädiktor für die Akzentähnlichkeit:

Analog zur Ermittlung des besten Einzelmaßes wurden auch hier die euklidischen Distanzen zwischen den *ratings* der Experten aus dem ersten Experiment zur melodischen Ähnlichkeit (Kapitel 4) und den verschiedenen *joint accent measures* über alle 84 Melodievergleichspaare berechnet. Die euklidischen Distanzen anhand der 32 Vergleichspaare aus dem Kontrollexperiment 3 dienten auch hier der Kontrolle.

Die 36 getesteten *joint accent measures* sind im Prinzip aus drei Komponenten konstruiert: Der spezifischen Kombination von mehreren Akzentregeln, der spezifischen Variante der Edit Distance und dem Vorhandensein des Regelpaars *shortphr+phrasbeg*. Die spezifischen Kombinationen der Akzentregeln lassen sich zweistufig hierarchisch beschreiben: Zum einen kann die Anzahl der kombinierten Akzentregeln unterschieden werden (4, 5, 6), zum anderen die Kombination der einzelnen Akzentregeln selbst. Da alle möglichen Ausprägungen dieser drei Komponenten vollständig untereinander kombiniert wurden, ist es möglich, in einem linearen Modell, die Faktorstufen zu ermitteln, die insgesamt die besten Ergebnisse der *joint accent measures* generieren.

Um die hierarchische Gliederung der spezifischen Akzentregelkombinationen zu berücksichtigen, wurden zwei getrennte lineare dreifaktorielle Modelle gerechnet, in denen die Anzahl der kombinierten Regeln bzw. die spezifische Variante als Faktor Eingang fanden. Das erste dreifaktorielle Modell nahm die euklidischen Distanz zu den *Expertenratings* als abhängige Variable und die drei Faktoren *Edvarian*, *Variant*, und *Phraseru* auf. Das Modell ist in der folgenden Tabelle zusammengefasst.

Tests der Zwischensubjekteffekte

Abhängige Variable: DISIM1AL

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	84,774 ^a	23	3,686	85,299	,000
Konstanter Term	13551,377	1	13551,377	313612,2	,000
EDVARIAN	72,625	5	14,525	336,142	,000
PHRASERU	,135	1	,135	3,119	,079
VARIANT	12,015	17	,707	16,356	,000
Fehler	8,296	192	,043		
Gesamt	13644,447	216			
Korrigierte Gesamtvariation	93,070	215			

a. R-Quadrat = ,911 (korrigiertes R-Quadrat = ,900)

Tabelle 5-10: Lineares Modell zur Erklärung der Distanz der Akzentähnlichkeitsmaße im Vergleich mit den *ratings* der Versuchspersonen aus Exp. 1 (Kapitel 4). (EDVARIAN = Variante der Edit Distance; PHRASERU = An- vs. Abwesenheit des Phrasenregelpaars *shortphr+phrasbeg*; VARIANT = spezifische Akzentregelkombination).

Das Modell weist eine sehr hohe Modellgüte gemessen an der aufgeklärten Varianz von 91,1% (korrigiertes R^2 : 0,9) auf.

Die Variante der Edit Distance (*Edvarian*) war ein hoch signifikanter Einflussfaktor. Das im Faktor *Phrasru* zusammengefasste Regelpaar, zeigte sich hier fast signifikant ($p = 0,079$) und die spezifische Kombination der Regeln im Faktor *Variant* erwies auch als hoch signifikanter Prädiktor.

Die besten Regelkombinationen waren jene, die in der obigen Tabelle mit den Nummern 58, 52, 56, 55, 63 und 64 wie in der folgenden Tabelle zu sehen ist.

Schätzungen

Abhängige Variable: DISIM1AL

VARIANT	Mittelwert	Standardfehler	95% Konfidenzintervall	
			Untergrenze	Obergrenze
41	7,987	,060	7,868	8,105
42	8,280	,060	8,161	8,398
43	8,134	,060	8,015	8,252
44	8,413	,060	8,294	8,531
45	7,922	,060	7,804	8,041
46	8,148	,060	8,029	8,266
51	8,020	,060	7,901	8,138
52	7,583	,060	7,464	7,701
53	7,888	,060	7,769	8,006
54	8,078	,060	7,960	8,196
55	7,680	,060	7,562	7,799
56	7,586	,060	7,467	7,704
57	7,800	,060	7,681	7,918
58	7,579	,060	7,461	7,698
61	7,973	,060	7,854	8,091
62	8,020	,060	7,901	8,138
63	7,711	,060	7,593	7,830
64	7,773	,060	7,655	7,892

Tabelle 5-11: Mittlere Distanzen der einzelnen Stufen des Faktors VARIANT (spezifische Akzentregelkombinationen) im Vergleich mit den *ratings* der Versuchspersonen aus Exp. 1 (Kapitel 4).

Diese Maße unterschieden sich untereinander nicht signifikant in einem Post-hoc-Vergleich nach Scheffé, wohl aber jedes von diesen von einer Reihe von anderen Maßen.

Für die Stufen des Faktors *Edvarian* ergaben sich in der folgenden Tabelle dargestellten Mittelwerte.

Bericht

DISIM1AL

EDVARIAN	Mittelwert	N	Standardabweichung
0	7,102693	36	,2592552
1	8,988872	36	,1921025
2	8,131009	36	,4224379
3	7,754581	36	,3509672
4	7,988203	36	,3291438
5	7,558982	36	,2647323
Insgesamt	7,920723	216	,6579404

Tabelle 5-12: Mittlere Distanzen der einzelnen Stufen des Faktors EDVARIAN (Varianten der Edit Distance) im Vergleich mit den *ratings* der Versuchspersonen aus Exp. 1 (Kapitel 4).

Die einfache Edit Distance ohne Kostenfunktion (Variante 0) war im Vergleich mit den *Expertenratings* signifikant besser als alle anderen Varianten. Die nächstbesten Varianten waren die beiden Maße, die die Akzentwerte in drei unterschiedliche Klassen einteilen, wobei die Variante 5 mit einer Kostenfunktion ausgestattet ist, die Variante 3 dagegen nicht. Diese beiden Varianten unterschieden sich untereinander nicht signifikant, waren aber signifikant besser als die drei übrigen Varianten (1, 2, 4).

Das zweite dreifaktorielle Modell, welches statt des Faktors *spezifische Regelkombination* (*Variant*) die Anzahl der kombinierten Regeln (*Nomeasur*) als Prädiktor aufnahm, ansonsten aber völlig analog gerechnet wurde, kam zu einem sehr ähnlichen Ergebnis: Auch hier war die Variante der Edit Distance ein hoch signifikanter Prädiktor ($p < 0,000$) und die Phrasenregeln nicht signifikant ($p = 0,216$). Die Anzahl der kombinierten Maße war zusätzlich eine hoch signifikante Einflussgröße ($p < 0,000$). Insgesamtklärte das Modell mit 85,5% (korrigiertes R^2 : 0,827) jedoch etwas weniger Varianz auf als das obige. Dies ist nicht verwunderlich, da im Faktor Anzahl der kombinierten Regeln (*Nomeasur*) ja weniger Information steckt als in der spezifischen Regelkombination.

Die Mittelwerte der drei Gruppen (4er-, 5er- und 6er-Maße) sind der folgenden Tabelle zu entnehmen.

Bericht

DISIM1AL			
NOMEASUR	Mittelwert	N	Standardabweichung
4	8,147149	72	,5827531
5	7,776615	96	,6691860
6	7,869300	48	,6621186
Insgesamt	7,920723	216	,6579404

Tabelle 5-13: Mittlere Distanzen der einzelnen Stufen des Faktors NOMEASUR (Anzahl der kombinierten Akzentregeln) im Vergleich mit den *ratings* der Versuchspersonen aus Exp. 1 (Kapitel 4).

Nach einem Scheffé-Test war nur der Mittelwert der 4er-Maße jeweils signifikant schlechter als der der 5er und 6er-Maße. Die beiden letzteren unterschieden sich dagegen nicht signifikant.

Zusammenfassen lassen sich die Ergebnisse dieses Modells dahingehend, dass die besten Maße die einfache Edit Distance ohne Kosten verwenden, aus Kombinationen von fünf oder sechs Regel bestehen und sich unter den mit den Nummern 58, 52, 56, 55, 63 oder 64

benannten spezifischen Regelkombinationen finden. Hinsichtlich des Beitrags des Phrasenregelpaars lässt sich keine Aussage treffen

Im direkten Vergleich der 216 *joint accent measures* mit den *Expertenratings* fanden sich unter den besten zehn Maßen denn auch acht, auf die die Beschreibung des Modells zutrifft. Dies spricht für die Vorhersagequalität des Modells. Die folgende Tabelle führt die zehn besten *joint accent measures* mit ihren Distanzen zu den *Expertenratings* auf. Im Suffix steht die erste Ziffer für die Variante der Edit Distance (wenn kein 3. Zahlzeichen, dann Variante 0) und die nächsten beiden Ziffern für die spezifische Regelkombination. Die Buchstaben ‚ph‘ bedeuten die Hinzunahme des Phrasenregelpaars:

JOINT52	6,5545
JOINT56	6,6288
JOINT56ph	6,709
JOINT52ph	6,7276
JOINT61	6,8234
JOINT63	6,8485
JOINT51	6,9005
JOINT63ph	6,916
JOINT55	6,9398
JOINT525	6,9411
JOINT58	6,9762

Tabelle 5-14: *Joint accent measures* mit geringster Distanz zu den *ratings* der Versuchspersonen aus Exp. 1 (Kapitel 4).

Das beste *joint accent measure* (*joint 52*) besteht also aus den fünf Regeln *jumpaft4* (Akzent auf Note nach Sprung von mindestens großer Terz), *pextrmf* (Akzent auf Konturumkehr in Müllensiefen-Frieler-Definition), *long2pr* (Akzent auf Ton mit Dauer >2xVorgängerdauer), *phrasend* (Akzent auf Phrasenende) und *triad* (Akzent auf Dreiklangston des zugrundeliegenden Akkordes); es verwendet die einfache Edit Distance ohne Kosten und enthält nicht das Phrasenregelpaar (kein Akzent auf Phrasenbeginn und auf 2. Ton von zweitönigen Phrasen).

Dass dieses optimierte *joint accent measure* auch absolut gesehen eine Erklärungskraft hinsichtlich der Ähnlichkeit der Melodievarianten besitzt, zeigt ein Vergleich mit den 35 getesteten Einzelmaßen des ersten Ähnlichkeitsexperimentes. *Joint52* stünde dort an achter Stelle gemessen an der Distanz zu den *Expertenratings*. Für ein Ähnlichkeitsmaß, welches nur Akzente, aber nicht den konkreten Tonhöhen- oder Dauernverlauf betrachtet, scheint dies ein erstaunlich gutes Ergebnis.

Zur Kontrolle wurde auf den Daten (32 Melodie-Variantenpaare) des zweiten Kontrollexperimentes der *Ähnlichkeitsratings* ein analoges Modell mit den selben Prädiktoren gerechnet, das ein fast gleiches Ergebnis – allerdings mit etwas schlechteren Modellgüteindikatoren (Modell mit *Variant* als Prädiktor: 81% der Varianz aufgeklärt, R^2 : 0,787) – erzielt: Variante der Edit Distance und spezifische Regelkombination bzw. Anzahl der kombinierten Maße sind hoch signifikante Einflussfaktoren. Zudem ist auch die An-/Abwesenheit des Phrasenregelpaars signifikant, mit besseren Ergebnisse bei Abwesenheit. Wie im obigen Modell ist die einfache Edit Distance ohne Kosten die effektivste. Die Regelkombinationen 54, 55, 63, 58, 56 und 64 waren die, mit der geringsten Distanz; und schließlich waren auch wieder die 4er-Kombinationen signifikant schlechter als die die 5er und 6er.

Die besten zehn Maße sind:

JOINT543	3,611
JOINT554	3,702
JOINT54	3,746
JOINT52	3,753
JOINT58	3,79
JOINT63ph	3,814
JOINT55ph	3,819
JOINT544	3,837
JOINT52ph	3,85
JOINT634	3,874

Tabelle 5-15: *Joint accent measures* mit geringster Distanz zu den *ratings* der Versuchspersonen aus Exp. 3 (zweites Kontrollexperiment, Kapitel 4).

Es entsprechen hier immerhin noch sechs *joint accent measures* der durch die drei Modellfaktoren vorhergesagten optimalen Faktorkombination.

Unter den jeweils zehn besten Maßen beider Datensätze und entsprechend der optimalen Faktorkombination des Modells sind die vier Maße: *joint52*, *joint58*, *joint52ph* und *joint63ph*.

Als bestes *joint accent measure* wurde insgesamt *joint52* ausgewählt, da es die geringste Rangsumme aus den beiden Tabellen (1+4=5) unter diesen vier Maßen aufweist.

Joint52 soll also als Prädiktor für die Akzentähnlichkeit im folgenden verwendet werden und damit die Variablen *Manipulationsklasse* und *Manipulationstyp* ersetzen.

5.6.1.2 Die Prädiktorenliste

Insgesamt stehen für die statistische Modellierung 19 Prädiktoren zur Verfügung, die in den vorangegangenen Abschnitten schon ausführlich erläutert wurden und hier der Vollständigkeit und Übersichtlichkeit halber noch einmal tabellarisch mit den im folgenden verwendeten Kurznamen und dem zugehörigen Skalenniveau aufgelistet werden sollen.

Kurzname	Skalenniveau	Gemessene Größe
zmusicco	Intervall	Indexvariable des Musikkonsums der Versuchspersonen
zmusicac	Intervall	Indexvariable der musikalischen Aktivität der Vpn.
mansecen	Intervall (fehlende Werte, wenn Testmelodie nicht manipuliert)	Mittlere Position der Töne der Referenzmelodie, die in der Testmelodie manipuliert sind; in Sekunden vom Ende gemessen.
manoteph	Intervall (fehlende Werte, wenn Testmelodie nicht manipuliert)	Position der ersten manipulierten Note in Phrase; in Noten vom Phrasenbeginn gemessen.
manphmel	Intervall (fehlende Werte, wenn Testmelodie nicht manipuliert)	Position der ersten manipulierten Phrase in der Testmelodie; in Phrasen vom Melodiebeginn gemessen.
similar	Intervall	Melodische Ähnlichkeit (opti1) zwischen Referenz- und Testmelodie
noterest	Dichotom	Repetierte Struktur ist in Testmelodie enthalten
harmfunc	Dichotom	Änderung der harmonischen Funktion zwischen den Melodietönen von Referenz- und Testmelodie
changera	Dichotom	Änderung des Ambitus zwischen den Melodietönen von Referenz- und Testmelodie
manrhyth	Dichotom	Änderung des Rhythmus
maninter	Dichotom	Änderung der Intervallfolge
mancont	Dichotom	Änderung der Kontur
numnoend	Intervall	Anzahl Noten zwischen Referenzmelodie und Ende
nummeend	Intervall	Anzahl Melodien zwischen Referenzmelodie und Ende
secsend	Intervall	Anzahl Sekunden zwischen Referenzmelodie und Ende
pocomel	Dichotom	Position Referenzmelodie in Stück (1. Teil, 2. Teil)
durtemel	Intervall	Dauer Testmelodie in Sekunden
lentemel	Intervall	Länge Testmelodie in Noten
joint52	Intervall	Akzentähnlichkeit, gemessen nach oben beschriebener Regelkombination joint52

Tabelle 5-16: Liste der 19 für die statistische Modellierung verwendeten Prädiktoren.

5.6.2 Das lineare Modell

5.6.2.1 Variablenselektion

Insgesamt stehen für die Modellierung der *ratings* der Versuchspersonen die 19 unabhängigen Variablen zur Verfügung. Weder ist zu erwarten, dass alle Variablen einen sinnvol-

len Erklärungsbeitrag liefern, noch wäre ein Modell mit 19 Variablen besonders transparent oder gut interpretierbar. Zusätzlich besteht die Gefahr in Regressionsmodellen, dass bei der Aufnahme von irrelevanten Variablen Suppressionseffekte ein verzerrtes Modell liefern (vgl. Bortz, 1999, S. 442ff).

Aus diesen Gründen ist eine Variablenselektion vor der eigentlichen Schätzung der Modellparameter eines linearen Modells unablässig.

Der Schritt der Variablenselektion findet auf dem ersten Daten-Testset statt, für das anhand von Zufallszahlen 14 aus den insgesamt 42 Versuchspersonen mit ihren zugehörigen Daten ausgewählt wurden. Insgesamt stehen in diesem ersten Datenset 331 Beobachtungen zur Verfügung.

Die Variablenselektion besteht in der Bildung von multivariaten linearen Modellen, die zur Erklärung der *ratings* der Versuchspersonen die 19 Prädiktoren bzw. Untermengen davon verwenden. Das statistische Software-Paket *R* stellt zwei Routinen zur Verfügung, mit denen durch iterative Prozesse die besten linearen Modelle zur Vorhersage einer Kriteriumsvariablen anhand einer Vielzahl von Prädiktoren ermittelt werden können und die auf verschiedenen Methoden zur Parameterschätzung beruhen. Dies sind die Funktionen *stepAIC* und *regsubsets*.

stepAIC (mit `direction="backward"`) stützt sich auf die Maximum-Likelihood-Schätzung (vgl. Bortz, 1999, S. 98f) von linearen Modellen und geht (bei der Option `direction="backward"`) von einem vollständigen Modell mit allen Hauptfaktoren und durch den Benutzer definierten Interaktionen aus. Schrittweise wird je ein Prädiktor entfernt und der Zuwachs der Modellgüte geprüft. Das Modellgütekriterium ist das sogenannte *Akaike Information Criterion* (AIC), welches wie folgt definiert ist:

$$AIC(k) = -2L_{n,k} + 2k$$

Dabei steht $L_{n,k}$ für den maximierten Logarithmus der Likelihood-Funktion, k für die Anzahl der Parameter und n die Anzahl der Beobachtungen (Akaike, 1974). Ziel der *stepAIC*-Routine ist es, dieses Kriterium zu minimieren, wobei größere Modelle mit dem Faktor 2 durch den ‚Strafterm‘ $2k$ bestraft werden. Die Routine stoppt beim kleinsten Modell, aus dem keine weiteren Variablen mehr ausgeschlossen werden können, ohne dass sich das AIC wieder erhöht.

Für den sich anschließenden Vergleich der besten Modelle aus der Routinen *stepAIC* und *regsubsets* wurde angesichts der großen Zahl von Termen im vollständigen Modell und der

dazu im Verhältnis geringen Anzahl von Beobachten (hier $n=331$) nicht das AIC, sondern eine Variante davon, das sogenannte Bayes Information Criterion (BIC) verwendet, welches statt des Strafterms $2k$ den Term $k \log n$ verwendet, wobei n für die Anzahl der Beobachtungen steht. In diesem Fall ist der Strafterm 2,5198 ($=\log 331$).

Die Routine *regsubsets* schätzt die Parameter verschiedener Modelle mit dem Kleinst-Quadrate-Kriterium (vgl. Bortz, 1999, S. 97f) und verwendet als Kriterium für deren Modellgüte R^2 bzw. die Summe der quadrierten Fehler. Zwar lässt sich zeigen, dass für normalverteilte Daten das Kleinst-Quadrate-Kriterium und die Maximum-Likelihood-Schätzung äquivalente Ergebnisse liefern, beide Routinen können jedoch aufgrund der unterschiedlichen Implementierung gerade bei Betrachtung von Interaktionstermen dennoch unterschiedliche Ergebnisse liefern, weswegen die Ergebnisse von beiden Routinen hier dargestellt werden sollen.

Im hier verwendeten *forward*-Modus beginnt die Routine *regsubsets* mit dem besten Prädiktor und fügt schrittweise weitere bis zu einer vom Benutzer definierten Grenze hinzu. Als Grenze wurden hier 20 Prädiktoren vorgegeben, da es inhaltlich nicht plausibel erscheint, dass die Versuchspersonen in der Experimentalsituation mehr als 20 Einflussvariablen berücksichtigen. Gleichzeitig sollte auch, um instabile Modelle zu vermeiden, die Anzahl der Prädiktoren zur Zahl der Beobachten in einem ‚gesunden‘ Verhältnis von $n/k > 15$ stehen, was bei $n=332$ ein k von maximal 22 ergibt. Als Output liefert *regsubsets* für jedes k , mit $k \in \{1, \dots, 20\}$ ein optimales Modell im Sinne der Modellgütekriterien.

Beiden Routinen wurde das selbe vollständige Modell aller 19 Hauptfaktoren und aller 2-Wege-Interaktionen als Input gegeben. Ausnahme waren die drei Variablen, die sich auf die Position bzw. die Stärke der Manipulation der Testmelodie beziehen (*mansecsen*, *manoteph*, *manphme*). Da diese nur dann gültige Werte besitzen, wenn überhaupt eine Manipulation vorliegt, waren sie nur als hierarchische Interaktion innerhalb des Faktors *similar* definiert. Nur wenn *similar* einen Wert < 1 hat (in Wahrheit ist der Wert von *similar*, der für Identität steht, 0,8603, da dies der maximale Werte des Regressionsmodells von *optil* ist), sind diese Interaktionen definiert. Diese hierarchischen Interaktion sind in der folgenden Formel mit *%in%* bezeichnet.

Das vollständige lineare Ausgangsmodell mit der abhängigen Variable *ratdacom* (*ratings* der Versuchspersonen), den 16 Hauptfaktoren, 3 hierarchischen Interaktionen und den 120

Interaktionen (insgesamt also 139 Terme in der Modellgleichung) als Input der beiden Routinen zur Variablenselektion ist mit seiner vollständigen Modellformel im Anhang zu finden.

Nach Anwendung der Routinen hatte *stepAIC* ein optimales mit 24 Prädiktoren und *regsubsets* 20 optimale Modelle (eins für jedes k) ermittelt. Um das beste aus diesen 21 Modelle zu selektieren, wurden die Modelle mit dem BIC-Kriterium miteinander verglichen.

Die folgende Tabelle gibt die Freiheitsgrade, die zu jedem Modell gehörigen BIC-Werte und zusätzlich die Summe der quadrierten Residuen (RSS) wieder. Da bei den Modellen aus *regsubsets* mit weniger als 10 Prädiktoren die BIC-Werte wieder ansteigen, werden die kleineren Modelle in der Tabelle nicht aufgeführt.

Modell	Terme im Modell	BIC-Werte	RSS
regsubsetsbest20	20	703,3609	244,90
regsubsetsbest15	15	698,2520	253,385
regsubsetsbest14	14	697,4620	255,385
regsubsetsbest13	13	696,3019	256,969
regsubsetsbest12	12	695,8530	259,399
regsubsetsbest11	11	695,4522	261,910
regsubsetsbest10	10	697,0398	266,84
Stepaicbest2 (mit $k=2$)	37	726,1673	223,591
Stepaicbest2.5 (mit $k=2,5$)	24	708,4983	239,464

Tabelle 5-17: BIC-Werte und quadrierte Residuen (RSS) zu den besten selektierten linearen Modellen.

Der beste BIC-Wert gehört zum Modell mit elf Prädiktoren, das durch *regsubsets* ermittelt wurde. Natürlich sind die Modelle, die mehr Prädiktoren enthalten, in der Lage, einen größeren Teil der Varianz in den Daten zu erklären und so einen geringen Wert bei der Quadratsumme der Residuen auszuweisen. Eine Überprüfung von *regsubsets11* und *stepaicbest2* (geringste RSS) mit Hilfe einer ANOVA ergab jedoch einen F-Wert von 1,1996, der bei 26 Freiheitsgraden keinen signifikanten Unterschied ($p>0,242$) zwischen den Modellen ergab.

Als bestes Modell wurde aufgrund dieser Daten das Modell *regsubsetsbest11* gewählt, das die folgende Modellformel enthält:

Modellformel:

RATDACOM ~ JOINT52+SIMILAR*MANOTEPH+
ZMUSICCO*DURTEMEL+ZMUSICAC*NOTEREST+
ZMUSICAC*SECSSEND+ZMUSICAC*MANRHYTH+
DURTEMEL*JOINT52+DURTEMEL*NOTEREST+
SIMILAR*HARMFUNC+MANINTER*HARMFUNC+
NUMNOEND*MANCONT

Wie auf den ersten Blick schon zu erkennen ist, war die Variablenselektion in der Hinsicht effektiv, als dass ein übersichtliches Modell mit einem Haupteffekt und 10 Interaktionseffekten gefunden wurde, welches für die weiteren analytischen Schritte handhabbar ist.

Das Modell bezieht sowohl Einflussfaktoren ein, die die Struktur der Testmelodie im Vergleich mit dem Original beschreiben (*joint52* sowie Interaktion zwischen *similar* und *manoteph*), als auch solche, die den musikalischen Hintergrund der Versuchspersonen mit konkreten Eigenschaften der Testmelodien verknüpfen (z.B. *zmusicac*noterest* oder *zmusicco*durtemel*). Auf eine weitere Darstellung und Interpretation der Modellparameter und ihrer Koeffizienten soll hier verzichtet werden, da dieser Schritt ja ausschließlich der Selektion der Variablen diene. Eine Interpretation soll im nächsten Schritt anhand der Modellformel inklusive der geschätzten Parameterwerte vorgenommen werden.

Hinsichtlich der für lineare Modelle geforderten Normalverteilung der Residuen scheint es die Voraussetzungen zumindest näherungsweise zu erfüllen, wie die folgende Verteilungstabelle zeigt und auch die visuelle Inspektion Plots der Residuen gegen die vorhergesagten Werte ergab:

Minimum	1. Quartil	Median	3. Quartil	Maximum
-3,472690	-0,808139	0,003291	0,812170	2,643615

Tabelle 5-18: Verteilung der Residuen des besten Modells nach der Variablenselektion.

Dieses Modell besitzt allerdings ein R^2 von nur 0,2364, erklärt also nur ca. 24% der Streuung in den Daten. Dies kann darauf hinweisen, dass die Urteile der Versuchspersonen durch zufällig bzw. sehr komplexe individuelle Effekte beeinflusst werden bzw. das noch ein oder mehrere wichtige Faktoren vom Modell nicht berücksichtigt wurden.

Für dieses Modell sollen im nächsten Schritt die Parameter für die 11 Prädiktoren geschätzt werden.

5.6.2.2 Parameterschätzung

Die Parameterschätzung für das obige Modell wurde mit der Prozedur *Gemischte lineare Modelle* in SPSS ausgeführt, da sich hier sowohl die Modellterme genau festlegen lassen als auch unter Berücksichtigung eines Zufallsfaktors die Parameter der Modellterme geschätzt werden können. Es wurden die elf selektierten Prädiktoren als feste Effekte verwendet sowie die Versuchsperson (*subjno*) als zufälliger Effekt.

Da es für die Vorhersage der abhängigen Variable (*Ratingdaten* der Versuchspersonen) in der Prozedur *Lineare Gemischte Modelle* unbedingt notwendig ist, dass alle Prädiktorvariablen einen gültigen Wert besitzen, wurden zwei verschiedene Modelle geschätzt: Eines auf den Daten zu den manipulierten Melodien unter Beteiligung der Variable *manoteph*, die nur gültige Werte besitzt, wenn eine Manipulation vorliegt. Das zweite wurde auf den Daten zu den identischen Testmelodien geschätzt und berücksichtigt konsequenterweise die Variable *manoteph* nicht. Da die Variablen *harmfunc*, *manrhyth*, *mancont* und *maninter* in den zu den identischen Testmelodien zugehörigen Daten notwendiger den konstanten Werte 0 aufweisen – denn bei identischen Melodien kann die harmonische Funktion, der Rhythmus etc. ja nicht geändert sein – wurden auch die vier Interaktionsterme, die diese Variablen enthielten, für die Parameterschätzung aus dem Modell für die identischen Melodien entfernt. Zusätzlich wurde auch die Variable *joint52* als Haupteffekt nicht mit in das Modell aufgenommen, da sie bei identischen Melodien notwendigerweise einen konstanten Wert zeigen muss, der sich nur auf die Schätzung des konstanten Terms (Achsenabschnitt) auswirkt.

Die Gewichte der Modellparameter für das Modell mit Manipulationen sind in folgender Tabelle wiedergegeben:

Schätzungen fester Parameter^a

Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Signifikanz	Konfidenzintervall 95%	
						Untergrenze	Obergrenze
Konstanter Term	2,1283061	,3958431	207,057	5,377	,000	1,3479065	2,9087058
SIMILAR * MANOTEPH	-,0864552	,0432348	203,602	-2,000	,047	-,1717007	-,0012098
ZMUSICCO * DURTEMEL	,0017647	,0059992	14,247	,294	,773	-,0110814	,0146109
ZMUSICAC * SECSSEND	,0055247	,0076848	208,614	,719	,473	-,0096252	,0206746
JOINT52 * DURTEMEL	,0809932	,0396452	205,164	2,043	,042	,0028289	,1591575
JOINT52	,3166195	,8181193	200,568	,387	,699	-1,2965990	1,9298381
ZMUSICAC * NOTEREST	-,6189571	,2006463	47,638	-3,085	,003	-1,0224628	-,2154514
ZMUSICAC * MANRHYTH	-,0019516	,2535431	200,763	-,008	,994	-,5019007	,4979976
DURTEMEL * NOTEREST	,0095695	,0172555	209,797	,555	,580	-,0244469	,0435859
SIMILAR * HARMFUNC	-,3240418	,3854709	204,187	-,841	,402	-1,0840555	,4359719
HARMFUNC * MANINTER	-,7030659	,2801848	203,986	-2,509	,013	-1,2554956	-,1506362
NUMNOEND * MANCONT	-,0001440	,0042379	204,744	-,034	,973	-,0084995	,0082116

a. Abhängige Variable: complete rating data.

Tabelle 5-19: Geschätzte Parameter der Terme des linearen Modells für die Vorhersage der *Versuchspersonenratings* bei manipulierten Testmelodien.

Die meisten Gewichte lassen sich ohne größere Schwierigkeiten interpretieren. Als Beispiele sollen hier nur die Parameter mit signifikanten Gewichten ($p < 0,05$) interpretiert werden:

similar x manoteph: Die Position der Manipulation in der Phrase korreliert negativ mit dem Identitätseindruck der Versuchspersonen (je später eine manipulierte Note in der Phrase auftaucht, desto geringer der subjektive Identitätseindruck), und dieser Zusammenhang ist umso stärker, je ähnlicher (melodische Ähnlichkeit nach *similar*) sich die beiden Melodievarianten sind.

joint52 x durtemel: Wie erwartet korreliert die Akzentähnlichkeit positiv mit den *ratings* der Versuchspersonen: Je ähnlicher sich die Akzentstrukturen von Referenz- und Testmelodie sind, desto stärker wird die Testmelodie als identisch bewertet. Dieser Zusammenhang gilt umso stärker, je länger die Testmelodie ist.

zmusicac x noterest: Dieser Interaktionsterm geht mit einem deutlich negativen Gewicht in die Gleichung ein. Da *noterest* als dichotome Variable angibt, ob eine repetierte Struktur vorliegt oder nicht, lässt sich dieser Term dahingehend interpretieren, dass musikalisch weniger vorgebildete Personen dazu tendieren, die Identität von Referenz- und Testmelodie zu überschätzen, wenn eine repetierte Struktur vorliegt; oder umgekehrt: Die musikalisch Aktiveren erkennen besser die Unterschiede zwischen den beiden Melodien, auch wenn eine repetierte Struktur in den Melodien enthalten ist.

harmfunc x *maninter*: Den größten Betrag zeigt das Gewicht zu diesem Interaktionsterm. Ändert sich die harmonische Funktion durch eine Manipulation der Intervalle, so hat dies den stärksten negativen Einfluss auf die *ratings* der Versuchspersonen.

Die übrigen Modellterme, die ausnahmslos aus Interaktionen bestehen, lassen sich als Zusammenhang zweier aneinander gewichteter Variablen mit den *ratings* der Versuchspersonen verstehen.

Das zweite Modell für die Melodien ohne Manipulation wurde – wie oben beschrieben – ohne jene Terme geschätzt, die bei identischen Testmelodien fehlende oder konstante Werte (v.a. 0) aufweisen. Die geschätzten Parameter gibt die folgende Tabelle wieder.

Schätzungen fester Parameter^a

Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Signifikanz	Konfidenzintervall 95%	
						Untergrenze	Obergrenze
Konstanter Term	3,0180201	,6137555	105,714	4,917	,000	1,8011521	4,2348881
ZMUSICCO * DURTEMEL	-,0003270	,0106160	13,781	-,031	,976	-,0231301	,0224762
ZMUSICAC * SECSEND	,0136110	,0076073	105,999	1,789	,076	-,0014712	,0286931
ZMUSICAC * NOTEREST	,3081842	,3849848	104,929	,801	,425	-,4551756	1,0715440
DURTEMEL * JOINT52	,1063277	,0483554	97,121	2,199	,030	,0103571	,2022982
DURTEMEL * NOTEREST	-,0245494	,0211416	98,426	-1,161	,248	-,0665019	,0174030

a. Abhängige Variable: complete rating data.

Tabelle 5-20: Geschätzte Parameter der Terme des linearen Modells für die Vorhersage der *Versuchspersonenratings* bei nichtmanipulierten Testmelodien.

Die beiden Terme mit den niedrigsten Irrtumswahrscheinlichkeiten dieses Modells sollen hier interpretiert werden:

zmusicac x *secsend*: die musikalisch Aktiveren vergeben höhere Werte für die identischen Melodien; dieser Zusammenhang ist umso stärker, je mehr Zeit zwischen der Referenzmelodie im Stück und der Testmelodie vergangen ist.

durtemel x *joint52*: Die Dauer der Testmelodie steht in einem positiven Zusammenhang mit den Urteilen zu den identischen Testmelodien. Da *joint52* bei den identischen Melodien immer den Wert 1 haben muss, ist hier eine Interpretation dieser Variable nicht notwendig. Sie ist nur aus ‚technischen‘ Gründen in diesem Interaktionsterm verblieben.

Die Güte dieses Modells (genauer: dieser beiden Modelle) soll erst im Abschnitt Modellevaluation detailliert untersucht werden. Das Modell scheint aber in jedem Fall die Vor-

aussetzung der näherungsweisen Normalverteilung der Residuen zu erfüllen, wie die folgende Grafik zeigt:

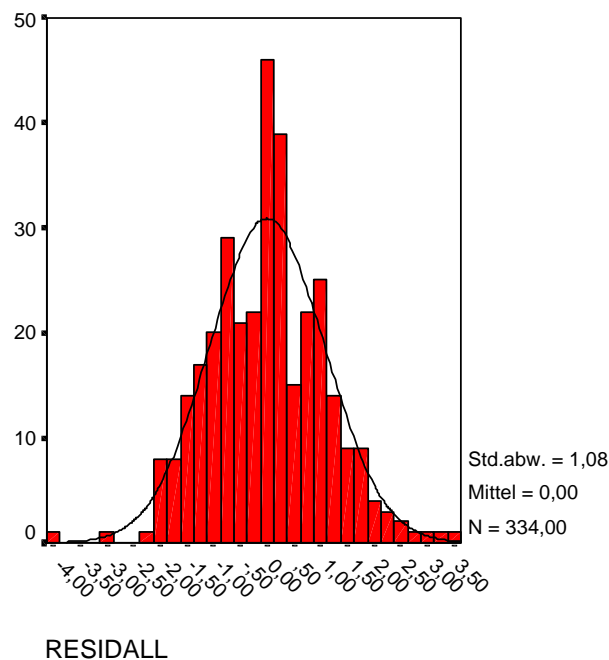


Abb. 5-7: Häufigkeitsverteilung der Residuen des linearen Modells nach der Parameterschätzung.

Dieses Modell ist nicht das einzige oder best mögliche im linearen Ansatz. Innerhalb der Operationalisierungen, der Variablenselektion sowie der Parameterschätzung wurden mehrere pragmatische oder heuristische Entscheidungen getroffen, die für die endgültige Gestalt des Modells mitverantwortlich sind. Einige dieser Entscheidungen hätten mit guten Argumenten auch anders getroffen werden können. Vielfach wurde jedoch eine Variante gewählt, die die Komplexität des Modells in einem beherrschbaren Rahmen belässt und die Arbeitsaufwand zumindest überschaubar werden ließ. Kurz zusammengefasst sind diese Entscheidungen die folgenden gewesen:

- Beschränkung auf intervallskalierte Kovariablen und dichotome Variablen.
- Beschränkung auf Interaktionsterme 2. Ordnung und Haupteffekte
- Modellierung der Interaktionen durch Multiplikation.
- Verwendung der *forward*-Methode der *regsubsets*-Routine.
- Verwendung der *backward*-Methode der *stepAIC*-Routine
- Konstruktion von zwei Modellen für manipulierte und nicht-manipulierte Testmelodien.

Diese heuristischen Entscheidungen werden beim Vergleich des linearen Modells mit den im folgenden beschriebenen Baum-Modellen zu berücksichtigen sein.

5.6.3 Die Baummodelle

5.6.3.1 Methode der Baummodelle

Sogenannte Klassifizierungs- und Regressionsbäume (*Classification and Regression Trees*) gehören mittlerweile zum Standardrepertoire in dem Bereich der Informatik, der sich mit dem Maschinellen (machine learning) befasst. Ihre Wurzeln haben diese Techniken in der automatischen Entdeckung von Interaktionen in linearen Modellen und in der Akquise von Wissen in sogenannten Expertensystemen, z.B. in medizinischen Anwendungen. Es hat sich mittlerweile ein Kanon von mehreren Algorithmen etabliert, die in verschiedenen Software-Paketen implementiert sind. Die wichtigsten Familien von Algorithmen sind die AID-Familie (zurückgehend auf Morgan & Sonquist (1963); Algorithmen: AID, CHAID, THAID etc.), die TDITD-Familie (frühe Veröffentlichungen von Hunt, Marin & Stone (1966) sowie Quinlan (1979); Algorithmen: CLS, ID3, C4.5 etc.) und die CART-Familie (v.a. Breiman et al. (1984); Algorithmen: CART sowie die Routinen *tree* und *rpart* des Softwarepaketes *R*).

Von diesen Modellen ist der CART-Algorithmus besonders gut für das vorliegende Problem geeignet, da er sowohl eine Möglichkeit zur Klassifizierung von Zuständen einer abhängigen Variable auf nominalem Niveau wie auch zur Regression einer kontinuierlichen abhängigen Variable auf mehrere unabhängige Variablen bietet.

Die mathematische Herleitung des CART-Algorithmus ist ausführlich in Breiman et al. (1984) dokumentiert. Seine Implementierung in der Routine *rpart* (= „recursive partitioning“) der statistischen Software *R* ist detailliert in Therneau & Atkinson (1997) beschrieben. Da Klassifizierungs- und Regressionsbäume jedoch bisher nicht zum Standard in der empirischen Musikforschung gehören, sollen in aller Kürze das Prinzip und die Grundüberlegungen dieser Technik geschildert werden, wobei so wenig Mathematik wie gerade eben nötig verwendet wird.

Klassifizierungs- und Regressionsbäume teilen Beobachtungen oder Fälle in Kategorien entlang einer unabhängigen Variable y ein, und zwar aufgrund der Ausprägungen auf einem Set von unabhängigen Variablen X . Ziel ist die Konstruktion eines stabilen Modells, das aufgrund von festen Regeln die Vorhersage der Ausprägungen auf der Variable y bei zukünftigen Fällen ermöglicht.

Dabei ist der Prozess der Konstruktion des Baumes („growing of the tree“) von der Evaluation seiner Stabilität zu unterscheiden. Baummethoden gehören zum Bereich des sogenannten überwachten Lernens („supervised learning“), d.h. sie lernen oder wachsen anhand von vollständigen Datensätzen mit gegebenem X und y (Lerndatensätzen), und können danach zur Vorhersage eines y bei gegebenem X eingesetzt werden. Da sie die Einteilungen der Beobachtungen – im Gegensatz zur linearen Diskriminanzanalyse – in iterativen und hierarchischen Prozessen vornehmen, lassen sich die in den Daten gefundenen Strukturen als hierarchische Bäume visualisieren. Die Bäume besitzen Knoten von denen aus sich Äste verzweigen. Die Knoten stellen Teilmengen des Gesamtdatensatzes dar und die Zweige führen von einer Teilmenge zu den Unterteilmengen. Der erste Knoten wird Wurzel des Baumes genannt und enthält alle Beobachtungen eines Datensatzes. Der CART-Algorithmus erlaubt nur Teilungen („splits“) in je zwei Äste pro Knoten. Da dies jedoch iterativ ausgeführt werden kann, stellt dieses Prinzip keine Einschränkung der Vorhersagegenauigkeit dar.

Die Konstruktion eines Klassifizierungs- oder Regressionsbaumes beschränkt sich im Prinzip auf drei Arten von Entscheidungen (vgl. Breimann et al., 1984, S. 22):

1. Wann und wie soll ein *split* vorgenommen werden, d.h. eine Gruppe von Fällen in einem Knoten aufgeteilt werden?
2. Wann ist ein Knoten ein Endknoten, der nicht weiter geteilt werden sollte bzw. wie wird die richtige Größe des Baums bestimmt?
3. Wie sollen die Endknoten benannt werden?

Die Voraussetzung für die beiden ersten Entscheidungen ist ein Maß für die Klassifizierungs- oder Regressionsgüte der Aufteilung von Knoten. Wird die Klassifizierung des Datensatzes insgesamt besser, wenn ein Knoten in zwei Äste geteilt wird und um wie viel wird die Gesamtklassifizierung besser?

Für Klassifizierung und Regression existieren unterschiedliche Güte- oder Reinheitsmaße. Ein weitverbreitetes Maß, das auch das default-Maß zur Klassifizierung in der *rpart*-Routine ist (bei nicht-Berücksichtigung einer Kostenfunktion), ist der Gini-Index. Er ist folgendermaßen definiert (vgl. Breiman et al., 1984, S. 38, 103f):

$$I(t) = \sum_{j \neq i} p(i | t) p(j | t)$$

$I(t)$ steht für die Unreinheit (impurity) der Beobachtungen in einem Knoten und ist die Bezeichnung des Indexes.

$p(i | t)$ ist Wahrscheinlichkeit dafür, dass ein zufällig ausgewähltes Objekt t der Klasse i zugeordnet ist.

$p(j | t)$ ist Wahrscheinlichkeit dafür, dass ein zufällig ausgewähltes Objekt t der Klasse j zugeordnet ist.

Die Wahrscheinlichkeiten $p(i)$ und $p(j)$ werden durch die relative Häufigkeiten der Fälle in einem Knoten geschätzt. Ziel der Klassifizierung entlang des Baumes ist es, den Gini-Index zu minimieren, d.h. die Knoten immer ‚reiner‘ werden zu lassen, oder anders formuliert: den Anteil einer Klasse von Fällen in einem Knoten zu erhöhen (und die Anteile der anderen Klassen gleichzeitig zu senken).

Es wird nun in jedem Knoten mit jeder abhängigen Variablen getestet, ob durch eine Trennung anhand der Werte der abhängigen Variablen der Gini-Index minimiert werden kann. Bei kategorialen unabhängigen Variablen wird jeder Wert gegen alle anderen Werte getestet, bei geordneten unabhängigen Variablen werden die Mittelpunkte zwischen benachbarten Werten überprüft, ob eine Trennung an diesem Punkt eine Minimierung des Indexes ermöglicht. Der CART-Algorithmus liefert eine erschöpfende Suche über alle denkbaren *split*-Möglichkeiten aller unabhängigen Variablen. Die größte Minimierung wird als *split* ausgewählt.

Im Prozess des Baumwachstums werden alle Knoten solange weitergeteilt, bis sie nur noch Beobachtungen mit einem Wert der y -Variablen enthalten.

Für Regressionsbäume, die eine geordnete abhängige Variable prognostizieren sollen, existiert ein anderes Kriterium für die Güte der Teilung in den Knoten. Dies ist ganz analog zu linearen Regression ein ‚ANOVA-Kriterium‘:

$SS_K - (SS_L + SS_R)$, wobei $SS_K = \sum (y_i - \bar{y})^2$, die Summe der Quadratabweichungen der beobachteten Werte in einem Knoten K vom Knotenmittelwert ist und SS_L und SS_R die Quadratsummeabweichungen im linken und rechten Subknoten sind. Dieses Kriterium soll maximiert werden. Das ANOVA-Kriterium schreibt also vor, einen Knoten dann zu tren-

nen, wenn die Summe der quadrierten Abweichungen in beiden Subknoten am geringsten ist. Die erschöpfende Suche nach dem besten *split*-Wert auf allen unabhängigen Variablen ist auch den Regressionsbäumen eigen.

Eine besondere Eigenschaft der Baummodelle ist ihr Umgang mit fehlenden Werten. Besitzt die unabhängige Variable, die zum besten *split* in einem Knoten führt, für einige Beobachtungen fehlende Werte, so können diese Beobachtungen nicht anhand dieser Variable klassifiziert werden. Statt dessen werden sie mit der besten ‚Surrogat-Variable‘ klassifiziert. Surrogat-Variablen sind solche unabhängige Variablen, die einen Großteil der Fälle gleich klassifizieren wie die beste unabhängige Variable. Diese Möglichkeit trotz fehlender Werte alle Beobachtungen klassifizieren bzw. vorhersagen zu können, ist ein großer Vorteil der Baum- gegenüber den linearen Modellen. Dort werden ja mit Absicht die Variablen in der Variablenselektion verworfen, die die selbe Information wie ein anderer Prädiktor tragen (d.h. eine hohe Korrelation aufweisen). Diese Information kann in einem Baummodell genutzt werden.

Die zweite der oben genannten Entscheidungen betraf die Bestimmung eines Knotens als Endknoten. Der Wachstumsprozess eines Baumes geht zunächst bis zu dem Punkt, an dem alle Endknoten Beobachtungen mit den selben Werten oder – im Regressionsfall – eine minimale Zahl von Beobachtungen (z.B. 5) enthalten. Es zeigt sich aber, dass diese vollständigen Bäume sehr instabil sind, d.h. bei leicht veränderten Daten eine andere Gestalt annehmen. Ziel ist es, einen stabilen Baum zu konstruieren, der auch für zukünftige Datensätze aus der selben Quelle eine Prognosequalität besitzt. Zusätzlich sind inhaltlich große Bäume mit vielen Verzweigungen oft schwer zu interpretieren. Die größere Komplexität eines Modells ist also ein Nachteil, der ‚bestraft‘ werden sollte. Ausgehend von einem vollständigen Baum ist also ein Prozess der Beschneidung („pruning“) notwendig, der das Baummodell auf eine stabile und nicht zu komplexe Größe stutzt. Es wird also ein Subbaum gesucht, der hinreichend gut klassifiziert, aber gleichzeitig nicht zu komplex ist, d.h. nicht zu viele Knoten besitzt.

Die Größe, an der die Güte der Subbäume gemessen wird, ist demnach ein kombiniertes Fehler-Komplexitätsmaß, das wie folgt notiert werden kann:

$$R_{\alpha}(K) = R(K) + \alpha |\tilde{K}|$$

$R(K)$ steht dabei für den Fehler (des jeweiligen Fehlermaßes), den ein Modell, bei einer gegebenen Anzahl von Knoten K macht und α ist eine Zahl >0 , die die Komplexität des Modells mit $|\tilde{K}|$ Endknoten bestraft. Die in CART (und in *rpart*) verwendete Methode ist die sogenannte *Crossvalidation* für die Schätzung des Fehlers und auch Schätzung von α . Vereinfacht lässt sich die *Crossvalidation* folgendermaßen beschreiben:

Es werden zunächst alle Beobachtungen in Subsets S_1, S_2, \dots, S_i zu etwa je 10 (in *rpart* verwendeter Wert) Beobachtungen eingeteilt. Für jede Kombination von Datensubsets der Form ‚alle Datensets ohne S_i ‘, werden dann folgende Schritte durchlaufen:

- Für jedes dieser Kombinationssubsets ohne S_i wird für jeden Wert von α , der einen Unterschied in $R_{\alpha}(K)$ macht, ein eigenes Baummodell gerechnet.
- Anhand jedes Modells wird der Wert der abhängigen Variablen für jede Beobachtung des Subsets S_i vorhergesagt.
- Der Vorhersagefehler jedes Modells wird anhand der beobachteten und vorhergesagten Werte in S_i berechnet (z.B. kleinste Quadrate oder Gini-Index).
- Über alle Subsets S_i werden die Vorhersagefehler für ein gegebenes α summiert. Das α mit dem geringsten Vorhersagefehler ist die beste Schätzung für den besten oder angemessensten Wert von α .
- Es wird ein Baummodell über alle Beobachtungen mit dem besten Wert von α gerechnet.
- Da sich in der Praxis ab einem gewissen Wert von α oft ein ‚Plateau‘ bei den Werten des Vorhersagefehlers einstellt, innerhalb dessen der Klassifizierungsfehler nicht mehr oder nur noch sehr geringfügig sinkt, verwenden CART und die *rpart*-Routine die sogenannte 1-SE-Regel. Es wird nicht das α mit dem absoluten Minimum des Vorhersagefehlers gewählt, sondern es werden alle α innerhalb des Intervalls von einem Standardfehler um das Minimum des Vorhersagefehlers betrachtet. Aus diesen Kandidaten wird das Modell mit der geringsten Komplexität, d.h. dem geringsten α -Wert, ausgewählt. In *rpart* ist der *default*-Wert für den minimalen ‚Komplexitätsstrafterm‘ (α), der noch einen Unterschied machen kann, bei 0,01 festgelegt.

Ist auf diese Weise ein α geschätzt und ein Baummodell mit einer bestimmten Anzahl von (End-)Knoten konstruiert worden, bleibt nur noch die dritte Entscheidung zu treffen, näm-

lich die Endknoten zu benennen. Dies ist vergleichsweise simpel: Im Fall von Klassifizierungsbäumen wird der Knoten nach dem Wert der meisten Beobachtungen von y benannt, die in diesem Knoten enthalten sind. Bei Regressionsbäumen wird als Name des Endknotens der Mittelwert aller Beobachtungen innerhalb des Knotens definiert.

Mit dieser leicht vereinfachten und knappen Darstellung der Baummodelle sollte die folgende Darstellung der Konstruktion zweier Modelle und deren Interpretation leicht zu verfolgen sein. Eine Unterteilung in Variablenselektion und Parameterschätzung ist nicht notwendig, da sich beides durch den Prozess des Baumwachstums und der Stutzung ergibt. Allerdings sollen sowohl ein Klassifizierungs- wie auch ein Regressionsbaum konstruiert werden. Denn bei der abhängigen Variable, der *ratings* von Identität/Verschiedenheit und Urteilssicherheit, handelt es sich genau genommen um eine Variable mit ordinaler Information und einem sehr begrenzten Wertebereich (6 verschiedene Werte). Dies stellt einen Grenzfall dar, bei dem es sowohl sinnvoll erscheint, möglichst viele Werte der abhängigen Variable nahe ihres wahren Wertes zu schätzen (Regression) als auch möglichst viele Werte genau einer Werteklasse zuzuordnen (Klassifizierung). Im Klassifizierungsansatz bleibt die ordinale Information, die in den Beobachtungen der abhängigen Variable vorhanden ist, unberücksichtigt.

5.6.3.2 Der Klassifizierungsbaum

Für den Klassifizierungs- wie auch später für den Regressionsbaum wurden genau jene 19 Prädiktoren verwendet, die auch in der Auswahl für das lineare Modell zur Verfügung standen.

Als einziger Parameter wurde abweichend von den in *rpart* gesetzten *default*-Werten die maximale Anzahl der Surrogat-Variablen auf 18 gesetzt, d.h. es konnte jede andere Prädiktorvariable für die beste Prädiktorvariable eines bestimmten *splits* in Frage kommen. Zusätzlich wurde der Komplexitätsparameter auf 0,005 gesetzt. Ansonsten wurde ein Standard-Klassifizierungsbaum gerechnet.

Das Baummodell wurde auf zwei Dritteln aller Daten entwickelt (insgesamt 665 gültige Beobachtungen), was exakt den Datensätzen für Variablenselektion und Parameterschätzung des linearen Modells entspricht.

Die vollständige Zusammenfassung des Klassifizierungsbaumes mit den Werten zu allen ersten *split*-Kriterien und allen möglichen Surrogat-Variablen findet sich im Anhang. Die Ergebnisse von Wachstums- und Stützungsprozess lassen sich hier folgendermaßen resümieren:

Der Klassifizierungsfehler im ‚Wurzelknoten‘ beträgt $477/665 = 0,7173$; also würde man 72% der Fälle falsch klassifizieren, wenn man alle Fälle der größten Klasse zuordnen würde.

Die Entwicklung des Vorhersage- bzw. *Crossvalidationfehlers* (*xerror*) erreicht ein Minimum bei zehn *splits* (elf Knoten), wie die folgende Tabelle zeigt. Wendet man die oben erwähnte 1-SE-Regel an, erfüllt das Modell mit sieben *splits* die beiden geforderten Voraussetzungen: Es besitzt einen *Crossvalidationfehler*, der innerhalb einer Standardfehlers zum Modell mit dem geringsten *Crossvalidationfehler* liegt ($0,8867925 + 0,02601050 > 0,9014675$), und es besitzt eine geringere Komplexität (größere Werte in der Spalte Komplexität zeigen eine geringere Komplexität des Modells an).

Komplexität	Anzahl splits	Relativer Fehler	Crossvalidation-Fehler	Crossvalidation-Standardfehler
0.104821803	0	1.0000000	1.0000000	0.02434495
0.018169113	1	0.8951782	0.9014675	0.02584276
0.011180992	4	0.8406709	0.8930818	0.02594027
0.008385744	7	0.8071279	0.9014675	0.02584276
0.006289308	10	0.7819706	0.8867925	0.02601050
0.005241090	14	0.7568134	0.8972746	0.02589207
0.005000000	18	0.7358491	0.9035639	0.02581769

Tabelle 5-21: Modellparameter von Klassifizierungsbäumen verschiedener Komplexität

Der relative Fehler gibt den Anteil der Fehlklassifizierungen wieder. Er ist auf den Wert im Wurzelknoten normiert. Um die absoluten Anteil der Fehlklassifizierungen zu erhalten muss man entsprechend den Wert des normierten relativen Fehlers mit dem Anteil der Fehlklassifizierungen in der Wurzel multiplizieren; in diesem Fall: $0,7173 \cdot 0,80713 = 0,57895$. Das Baummodell mit sieben *splits* klassifiziert also immer noch 58% der Fälle falsch.

Insgesamt wurde also ein Baum mit acht Endknoten selektiert. Unter den Endknoten befinden sich drei mit dem *y*-Wert 3, sonst ist für jeden *y*-Wert (1,2,4,5,6) ein Endknoten vorhanden.

Es wurden im gesamten Baum nur vier Variablen verwendet, nämlich *durtemel*, *similar*, *numnoend*, und *zmusicco*. Die Variable *zmusicco* für insgesamt drei *splits*.

Den besten Überblick über das Modell gibt jedoch die folgende grafische Darstellung:

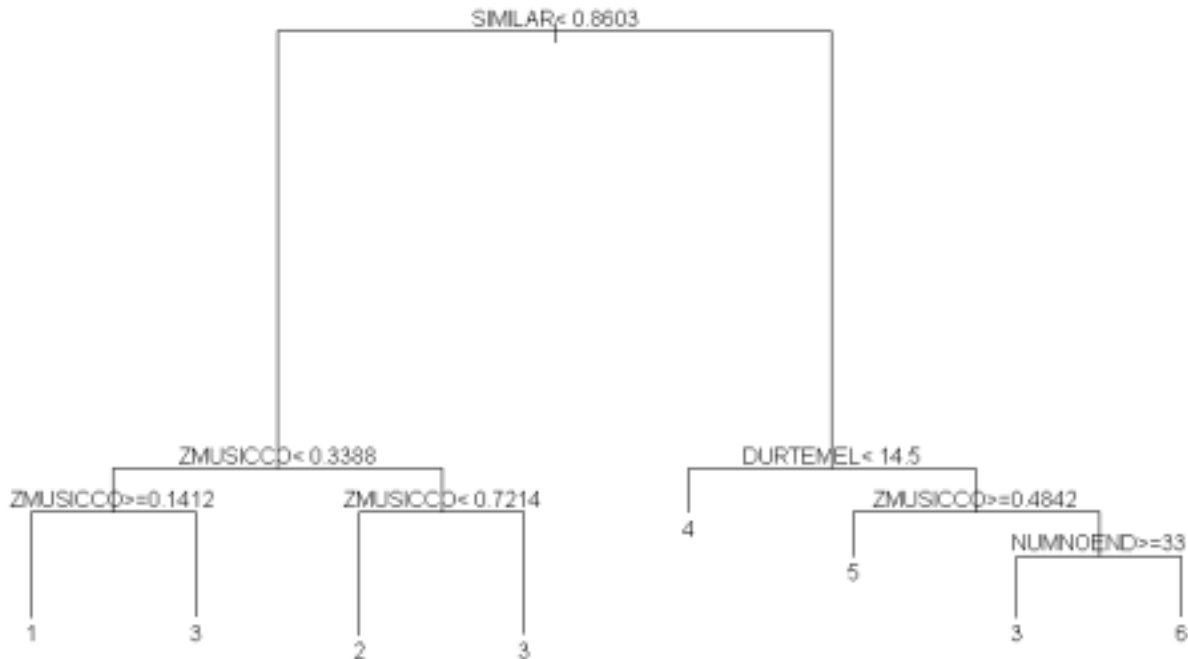


Abb. 5-8: Klassifizierungsbaum für die abhängige Variable *ratings* der Versuchspersonen (Werte der 6-stufigen *rating*-Skala).

Über jedem Knoten ist das jeweilige *split*-Kriterium angegeben. Bei Zutreffen des Kriteriums wird ein Fall nach links ‚geschickt‘, sonst nach rechts. Die Länge der Äste ist proportional zur Verbesserung im Reinheitsmaß (Gini-Index). Das wichtigste Kriterium ist also, ob die Melodien identisch sind oder nicht (*similar* < 0,8603). Bei manipulierten Melodien entscheidet dann nur noch der Musikkonsum über die Ähnlichkeitswerte. Die Versuchspersonen mit einem mittleren z-standardisierten Musikkonsum zwischen 0,1412 und 0,7214 vergeben richtige Ähnlichkeitsurteile mit einer größeren Sicherheit (*rating*-Werte 1 und 2). Die Versuchspersonen mit sehr hohem oder niedrigem Musikkonsum sind sich unsicherer in der Bewertung und vergeben eine 3 für die nicht-identisch Melodien. Nimmt man doch die ordinale Information zur Interpretation, so zeigt sich hier ein nicht-linearer Zusammenhang zwischen Musikkonsum und *rating*-Werten, der in einem linearen Modell nicht hätte aufgedeckt werden können. Handelt es sich um nicht-manipulierte Testmelodien, so ist zunächst die Dauer der Testmelodie in Sekunden entscheidend: Handelt es sich um eine kurze Melodie von weniger als 14,5 Sekunden, so bekommt sie den niedrigeren Wert 4 auf der Urteilssicherheitsskala. Ist es dagegen eine längere Melodie, so vergeben

die Versuchspersonen mit einem höheren Musikkonsum eher den Wert 5. Für die Versuchspersonen mit einem niedrigeren Musikkonsum ist dann noch die Anzahl der Noten entscheidend, die zwischen dem ersten Hören der Melodie im Stück und dem zweiten Hören in der Testphase liegen: Sind mehr als 32 Noten zu hören gewesen, vergeben sie den niedrigeren Wert 3, sind es weniger so bekommt die identische Melodie den Höchstwert 6. Natürlich sind 58% falsch klassifizierte Werte absolut gesehen kein besonders guter Indikator für das Modell. Betrachtet man aber seine Struktur, so macht es einen sehr plausiblen Eindruck: Mit nur einer Ausnahme sind die ‚richtigen‘ Werte hinsichtlich der Manipulation in den linken und rechten Hauptast aufgeteilt. Dies ist umso bemerkenswerter als das Modell ja keinerlei Ordnungsinformation über die abhängige Variable besaß. Eine genaue Evaluation des Modells soll erst durch den Vergleich der beobachteten Werte des dritten Teildatensatzes mit den vorhergesagten stattfinden. Allerdings ist das Modell relativ gut zu interpretieren. Am wichtigsten ist die Ähnlichkeitsrelation, die durch *similar* gemessen wird, danach ist die Höhe des Musikkonsums entscheidend und schließlich speziellen Eigenschaften der Testmelodie (Dauer und Abstand zu ihrer Referenz im Stück), die durch das Versuchsdesign bestimmt sind.

5.6.3.3 Der Regressionsbaum

Auch der Regressionsbaum wurde mit den *default*-Werten der *rpart*-Routine gerechnet, wieder mit Ausnahme der maximalen Zahl der Surrogat-Variablen von 18 und wieder mit einem Komplexitätsschwellwert von 0,005.

Die ausführliche Zusammenfassung der Splitkriterien findet sich im Anhang. Die wichtigsten Ergebnisse sind – analog dem obigen Schema – die folgenden:

Die Summe der quadrierten Abweichen beträgt im Wurzelknoten 1235,42, pro Beobachtung ist die Abweichung $1235,42/665 = 1,8578$. Die Entwicklung der *Crossvalidation*werte entlang der veränderten Komplexitätswerte zeigt ein Minimum bei fünf *splits*. Nach Prüfung durch die 1-SE-Regel auf das Modell mit fünf *splits*, d.h. sechs Endknoten.

Komplexität	Anzahl splits	Relativer Fehler	Crossvalidation-Fehler	Crossvalidation-Standardfehler
0.180054230	0	1.0000000	1.0039323	0.04666905
0.023792255	1	0.8199458	0.8524988	0.04261630
0.018845391	2	0.7961535	0.8599105	0.04322279
0.013024838	3	0.7773081	0.8323163	0.04280122
0.012898591	4	0.7642833	0.8291916	0.04417916
0.008928146	5	0.7513847	0.8223412	0.04380699
0.008106478	8	0.7246003	0.8330975	0.04469608

Tabelle 5-22: Modellparameter von Regressionsbäumen verschiedener Komplexität.

Der absoluten quadrierten Abweichungen sinken von durchschnittlich 1,8578 auf 1,395 für das Modell mit sechs Endknoten.

Für das Modell mit sechs Endknoten werden nur vier verschiedene Variablen benötigt: *similar*, *joint52*, *lentemel* und *zmusicac*; inhaltlich also eine identische und drei sehr ähnliche Größen, wie beim obigen Klassifizierungsmodell.

Die folgende Abbildung macht das Modell leicht verständlich. Sie zeigt für jeden Endknoten den Mittelwert der abhängigen Variable (*ratdacom*). Die Knoten, die keine Endknoten darstellen, sind die Äste mit der jeweiligen Teilungsregel beschriftet.

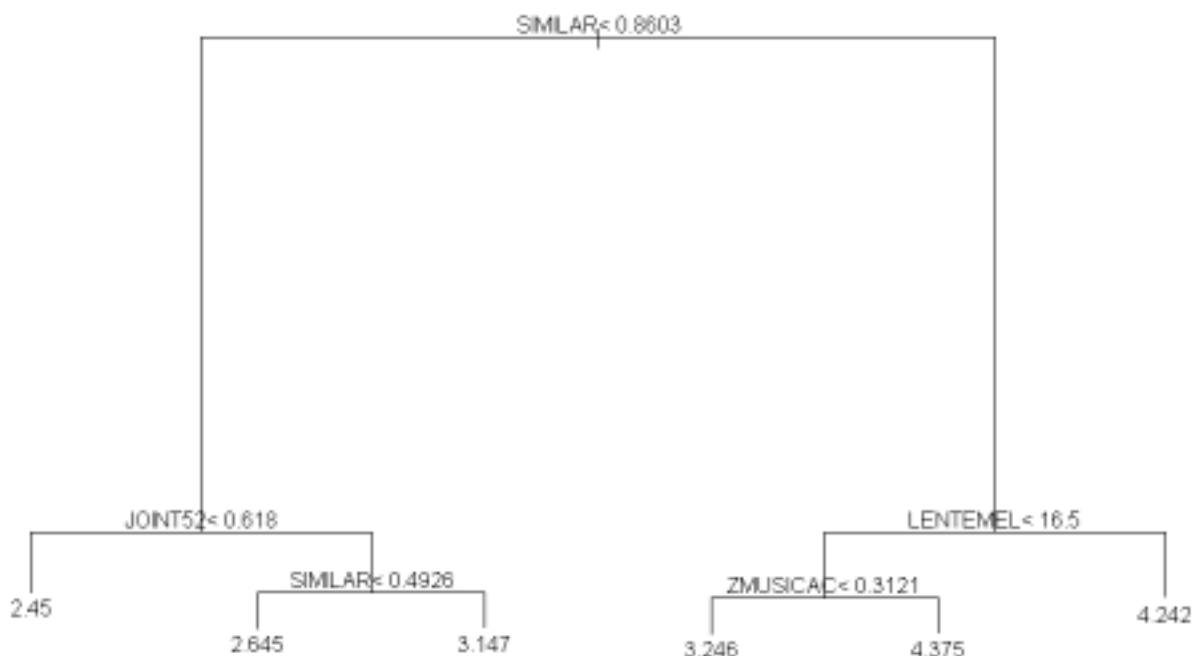


Abb. 5-9: Regressionsbaum für die abhängige Variable *ratdacom* (Werte der 6-stufigen *rating*-Skala).

Da zwei Endknoten weniger vorhanden sind, ist die Interpretation einfacher als die des Klassifizierungsbaumes: Den größten Fortschritt bringt die Teilung anhand der melodi-

schen Ähnlichkeit. Je geringer die Ähnlichkeitswerte, desto geringer auch die Gruppenmittelwerte in den Endknoten. An diesem Beispiel lässt sich gut erkennen, dass es sich wirklich um ein iteratives Verfahren handelt, welches den Wertebereich einer Variablen durch rekursiv ausgeführte *splits* in mehrere verschiedene Bereiche einteilen kann, denn die Variable *similar* wird nicht nur im Wurzelknoten zur Teilung verwendet, sondern auch für die Teilung in zwei Endknoten. Bei den identischen Testmelodien (rechter Ast) wird – fast wie im obigen Modell nach der Länge der Testmelodie aufgeteilt: Die kürzeren bekommen niedrigere Werte, die längeren höhere. Schließlich macht die musikalische Aktivität der Versuchspersonen bei den kürzeren identischen Testmelodien noch einen Unterschied. Die Versuchspersonen mit geringerer musikalischer Aktivität schätzen die kurzen Melodien eher unähnlich ein, die Versuchspersonen mit größerer Vorbildung dagegen deutlich eher identisch. Die Interpretation dieses Modells ist insgesamt sehr plausibel.

Der Nachteil dieses Regressionsmodells liegt in der Berechnung der Abweichungen vom Knotenmittelwert. Da die beobachteten Werte nur ‚zur Mitte der Skala hin‘ von den Extremwerten der Skala (1 und 6) abweichen können, ist es nicht erstaunlich, dass sich keine Gruppen mit den Mittelwerten im Bereich der Ränder der Skala gefunden haben. Rundet man die Gruppenmittelwerte auf ganze Zahlen, so erhält man eine Gruppe für den Wert 2, drei Gruppen für den Wert 3 und zwei Gruppen für den Wert 4. Extremwerte, d.h. die gerundeten Gruppenmittelwerte 1, 5 oder 6, können durch dieses Modell nicht vorhergesagt werden.

5.6.3.4 Das zweistufige Klassifizierungsmodell

Wie an der kurzen Beschreibung der beiden Baummodelle deutlich geworden ist, haben beide Modelle Schwierigkeiten mit den Eigenheiten der sechsstufigen Ordinalskala, die hier als abhängige Variable vorhergesagt werden soll. Das Klassifizierungsmodell kann die ordinale Information zwischen den Werten der abhängigen Variable nicht ausnutzen, d.h. zum Beispiel die Fehlklassifizierung eines Wertes in die Gruppe 6, der in Wahrheit in die Gruppe 2 gehört, ist für das Modell nicht falscher, als die Klassifizierung in die Gruppe 3.

Der Regressionsbaum berücksichtigt zwar die Distanzen zwischen den durch das Modell vorhergesagten Werten und den wahren Werten, hat aber durch die Begrenzung der Skala Schwierigkeiten, Werte an den Skalenenden vorherzusagen.

Aus diesem Grund soll in diesem Abschnitt ein zweistufiges Klassifizierungsmodell konstruiert werden. Das Modell nimmt im ersten Schritt nur eine dichotome Einteilung der

Werte vor, und zwar bez. der Entscheidung der Versuchspersonen, ob eine Testmelodie identisch (Werte 4-6) oder verschieden (Werte 1-3) zur Referenzmelodie gewesen ist. In einem zweiten Schritt wird dann die Urteilssicherheit innerhalb dieser beiden Gruppen vorhergesagt. Da sich die Urteilssicherheit auf drei unterschiedliche Werte beschränkt, sollte der Fehler, den das Klassifizierungsmodell aufgrund der Nicht-Beachtung der ordinalen Information macht, geringer sein, als bei einem Vorhersagebereich von sechs unterschiedlichen Kategorien.

Im ersten Schritt sollte also das Identitätsurteil der Versuchspersonen vorhergesagt werden, welches durch die durch Umkodierung geschaffene dichotome Variable *ident* ausgedrückt wird. Die Parameter für den ersten Klassifizierungsbaum wurden analog zu den beiden obigen Durchläufen gesetzt. Allerdings wurde ein Minimum von zwei *splits* vorausgesetzt, da erst dann ein zweistufiges Vorgehen sinnvoll erscheint. Die folgende Tabelle gibt die Entwicklung der Klassifizierungsfehler und der Komplexität wieder.

Komplexität	Anzahl splits	Relativer Fehler	Crossvalidation-Fehler	Crossvalidation-Standardfehler
0.357142857	0	1.0000000	1.0000000	0.04356146
0.018707483	1	0.6428571	0.6428571	0.03956182
0.006802721	3	0.6054422	0.6326531	0.03937006
0.005830904	4	0.5986395	0.6904762	0.04039344
0.005102041	16	0.5204082	0.6904762	0.04039344
0.005000000	18	0.5102041	0.6904762	0.04039344

Tabelle 5-23: Modellparameter von Klassifizierungsbäumen verschiedener Komplexität auf der ersten Stufe des zweistufigen Modells (Vorhersage: Identisch/Verschieden).

Das Modell, welches alle geforderten Bedingungen erfüllt (mindestens zwei *splits*, geringste Komplexität im Bereich von einer Standardabweichung um den geringsten *Crossvalidationfehler*), ist das Modell mit drei *splits*. Dieses Modell verwendet die Variablen *similar*, *secsend*, und *numnoend*. Der anfängliche Klassifizierungsfehler von $1 - (665 - 294)/665 = 0,442$ verbessert sich für das gewählte Modell auf 0,267. Das Modell sagt also knapp 27% der Antworten der Versuchspersonen auf der dichotomen Variablen falsch voraus. Grafisch ist es in der folgenden Abbildung repräsentiert.

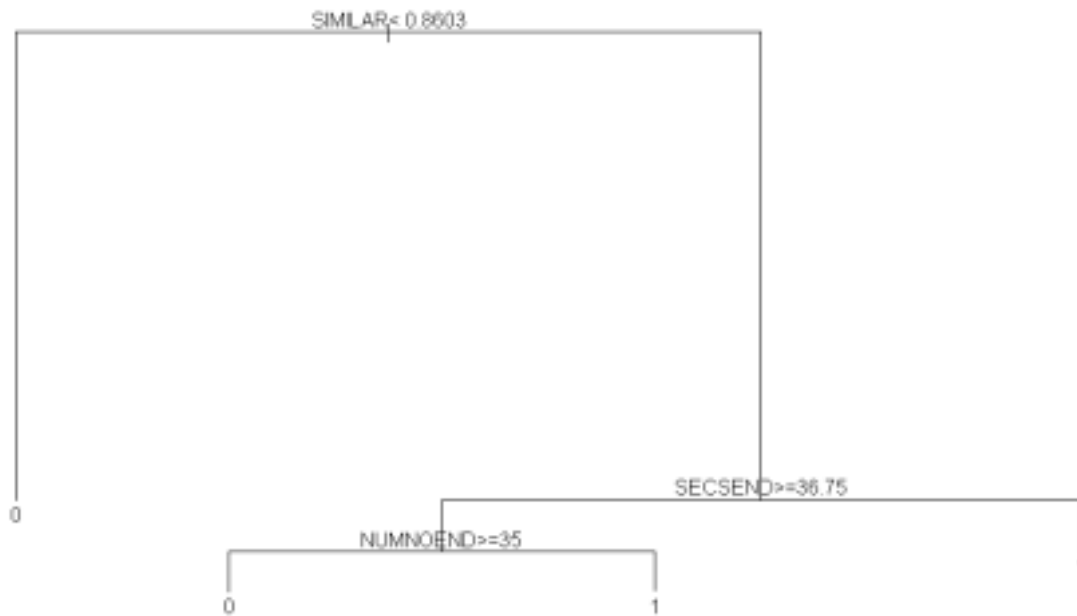


Abb. 5-10: Erste Stufe des zweistufigen Klassifizierungsbaum; Vorhersage des dichotomen Urteils Identisch/Verschieden.

Am wichtigsten zur Vorhersage des Identisch/Verschieden-Urteils der Versuchspersonen ist demnach die messbare Ähnlichkeit. Bei den Testmelodien, die zur Referenzmelodie identisch sind, spielt dann noch der Abstand zwischen Referenzmelodie im Stück und der Testmelodie eine Rolle, und zwar zuerst gemessen in Sekunden und danach in dazwischenliegenden Tönen: Sind weniger als 36,75 Sekunden vergangen, dann sagt das Modell voraus, dass die Versuchspersonen die Melodie richtigerweise als identisch erkennt. Sind 36,75 oder mehr Sekunden vergangen, dann spielt die Anzahl der Töne eine Rolle: Viele Töne bis zum Stückende (d.h. 35 oder mehr) lassen die Versuchspersonen offensichtlich die Referenzmelodie vergessen und generieren ein Verschieden-Urteil, weniger als 35 Töne führen dagegen zum einem richtigen Erkennen der identischen Testmelodie.

In der zweiten Stufe stehen nun diese durch das Modell vorhergesagten Werte des Identitätsurteils als zusätzlicher Prädiktor (*idpre*) zur Verfügung. Außerdem wurde ein neuer Prädiktor geschaffen, der als dichotome Variable die Übereinstimmung zwischen vorhergesagtem Identitätsurteil und tatsächlicher Identität von Referenz- und Testmelodie angibt. Die im zweiten Schritt vorherzusagende Größe ist die Urteilssicherheit der Versuchspersonen, die drei unterschiedliche Werte annehmen kann. Die Tabelle der Komplexität und den Klassifizierungsfehler

Komplexität	Anzahl splits	Relativer Fehler	Crossvalidation-Fehler	Crossvalidation-Standardfehler
0.0282392	0	1.00000	1.00000	0.042644
0.0265781	2	0.94352	0.97342	0.042533
0.0232558	3	0.91694	0.93355	0.042320
0.0199336	4	0.89369	0.92691	0.042278
0.0132890	6	0.85382	0.92027	0.042236
0.0099668	7	0.84053	0.91030	0.042168
0.0091362	13	0.76744	0.96013	0.042468

Tabelle 5-24: Modellparameter von Klassifizierungsbäumen verschiedener Komplexität auf der zweiten Stufe des zweistufigen Modells (Vorhersage der Urteilssicherheit).

Nach der 1-SE-Regel ist das Modell mit drei *splits* zu wählen. Dieses Modell verwendet nur die Variable *zmusicco*, die den Musikkonsum der Versuchspersonen wiedergibt. Nach diesem Modell hängt also die Neigung, Extremwerte auf der Skala zu vergeben, nur vom Musikkonsum der Versuchspersonen ab. Die Variable *zmusicco* zeigt allerdings einen nicht-linearen Zusammenhang mit der Urteilssicherheit, wie auch in der folgenden Grafik zu erkennen ist.



Abb. 5-11: Zweite des zweistufigen Klassifizierungsbaums; Vorhersage der Urteilssicherheit.

Personen, die einen sehr geringen (kleiner als Indexwert 0,14) oder sehr hohen (größer/gleich 0,72 auf der Indexskala) Musikkonsum über die *items* des Fragebogens angaben, neigen dazu, bei der Urteilssicherheit geringe Werte (Wert 1) anzugeben. Personen mit mittlerem Musikkonsum wählen dagegen eher auch die höheren Werte (Werte 2 und 3) auf der Skala zur Urteilssicherheit.

5.6.4 Evaluation der Modelle

Die Evaluation der Modelle wurde auf dem dritten, bisher nicht verwendeten, Datensatz vorgenommen. Die Daten stammen von 14 neuen Versuchspersonen und ein Drittel der zwölf Teststücke sind ebenfalls nicht in den bisher verwendeten Datensets zur Variablen-selektion und Parameterschätzung vorhanden gewesen. Die Modellevaluation findet also quasi auf Daten eines neuen Experimentes statt, welches unter den gleichen Bedingungen durchgeführt wurde. Replikationen des Experiments sollten zur selben Modellgüte führen. Dies kann als methodologischer Fortschritt gegenüber experimentellen Ergebnissen gesehen werden, die an einem einzigen Datensatz konstruiert und validiert wurden, wie dies in der Musikpsychologie eher die Regel ist.

Zur Bestimmung der Modellgüte bieten sich drei unterschiedliche Maße an, da im vorhin- ein nicht klar ist, wie genau die ‚Vorhersagegüte‘ des Modells definiert sein soll.

Das erste Maß, zählt die richtig vorhergesagten Werte der abhängigen Variablen. Bei den beiden Modellen, die sich nicht auf die Werte ganzer Zahlen beschränken, werden die vorhergesagten Werte zunächst auf die nächste ganze Zahl gerundet. Dann wird das Kriterium

R_1 wie folgt berechnet: $R_1 = \frac{1}{n} \sum_{l_i=m_i; i=1}^n 1$, mit $l_i \in L = \{l_1, l_2, \dots, l_n\}$ für die Menge der vorherge-

sagten und danach gerundeten Werte und $m_i \in M = \{m_1, m_2, \dots, m_n\}$ für die beobachteten Werte.

Das zweite Maß zählt nur die richtigen Vorhersagen des identisch/verschieden-Urteils. Dazu werden die beobachtete Variable und die vorhergesagten Werte dichotomisiert

(Werte 1-3 \Rightarrow 0; Werte 4-6 \Rightarrow 1). $R_2 = \frac{1}{n} \sum_{c_i=d_i; i=1}^n 1$, mit $c_i \in C$ für die Menge der vorhergesag-

ten und am Werte 3,5 dichotomisierten Werten und $d_i \in D$ für die Menge der auf die gleiche Weise dichotomisierten beobachteten Werte.

Das dritte Maß zählt die relative Distanz zwischen den beobachteten und den vorhergesagten Werten, es lässt sich als relativer mittlerer quadrierter Fehler bezeichnen. Um für vergleichbare Werte der Modellgüte auch für zukünftige Experimente zu schaffen, werden die Modellgütewerte alle an der Distanz der beobachteten Werte vom Mittelwert des Test-

samples (Datensets 1+2) normiert. So hängt die Modellgüte auch nicht mehr von der für die abhängige Variable verwendeten Skala ab, wie Breimann et al. (1984, S. 224) erläutern.

$R_3 = \frac{\frac{1}{n} \sqrt{\sum (y_i - \hat{y}_i)^2}}{\frac{1}{n} \sqrt{\sum (y_i - \bar{y})^2}}$, mit y_i für die beobachteten Werte des Testsample (Datenset 3)

und \hat{y}_i für die durch das Modell vorhergesagten Werte sowie \bar{y} für den Mittelwert der beobachteten Werte der Datensets 1 und 2, dieser hat den Wert 3,31. R_3 hat immer positive Werte. Modelle, die bessere Vorhersagen machen als der Mittelwert bekommen Werte <1 , schlechtere Modelle Werte >1 .

5.6.4.1 Ergebnisse des linearen Modells

Von den 330 gültigen Beobachtungen im Datenset 3 wurden 106 richtig vorhergesagt. Dies entspricht einer Trefferquote von $R_1 = 0,321$ oder 32,1%. Der häufigste beobachtete Wert in den Datensets 1 und 2 war der Wert 3. Der Wert 3 hat im Datenset einen Anteil von 24,8%. Sich für den häufigsten Wert zu entscheiden, bringt also kein besseres Ergebnis als das lineare Modell hinsichtlich des Maßes R_1 .

Für die Vorhersage der Kategorien identisch/verschieden erreichte das lineare Modell 213 richtige Vorhersagen bei 330 Fällen oder $R_2 = 0,645$, also 64,5% richtige Vorhersagen. Dieser Wert ist deutlich besser als die Ratewahrscheinlichkeit ohne Vorwissen (50%) und auch als die Entscheidung für die größere Gruppe (Werte 1-3) aus den Datensets 1 und 2, die in 48,2% der Fälle eine Fehlentscheidung gewesen wäre. Für die Vorhersage, ob die Versuchspersonen die Testmelodie für identische oder verschieden halten, ist das lineare Modell also besser geeignet. Der Zusammenhang zwischen den dichotomen beobachteten und vorhergesagten Werten erweist sich in einem χ^2 -Test als hoch signifikant ($\chi^2 = 15,257$, $df = 1$, $p < 0,001$).

Für R_3 ergibt sich ein Wert von 1,002. Das lineare Modell macht also Vorhersagen, die in der Summe genauso stark von den beobachteten Werten abweichen wie es der Mittelwert der Datensets 1 und 2 tut.

Zusammengenommen lässt sich sagen, dass das lineare Modell für genaue Vorhersagen der *ratings* der Versuchspersonen nicht besonders geeignet erscheint. Zu groß sind die Abweichungen der vorhergesagten von den beobachteten Werten. Für die Entscheidung, ob die Vpn. die Testmelodie als identisch oder verschieden beurteilt, erreicht das Modell eine deutlich bessere Vorhersagegüte von 2/3. Das Modell modelliert also die Gedächtnisleistung im Wiedererkennensexperiment deutlich besser als die Urteilssicherheit.

Die vorhergesagten Werte korrelieren mit den beobachteten hochsignifikant ($p < 0,001$, einseitiger Test) positiv mit einem Werte von $r = 0,396$ nach Pearson-Bravais.

5.6.4.2 Ergebnisse des Klassifizierungsbaumes

Von den 330 Beobachtungen sagt der Klassifizierungsbaum 32,1% (106 Beobachtungen) korrekt vorher ($R_1 = 0,321$). Dies ist leicht besser als die Vorhersagen des linearen Modells.

Die Entscheidung der Versuchspersonen hinsichtlich verschieden/identisch sagt der Klassifizierungsbaum genauso gut vorher das lineare Modell: 216 der 330 Beobachtungen sind richtig vorhergesagt, $R_2 = 0,655$.

Auch hier ergab sich ein hochsignifikanter Zusammenhang zwischen den Entscheidungen der Versuchspersonen und den vorhergesagten Werten ($\chi^2 = 31,765$, $df = 1$, $p < 0,001$)

Für die relative mittlere quadrierte Abweichung der durch den Klassifizierungsbaum vorhergesagten Werte von den beobachteten Werte ergibt sich ein Wert von $R_3 = 1,062$, der anzeigt, dass die Vorhersage leicht schlechter ist als die des Mittelwerts der Datensätze 1 und 2. Die Korrelation nach Pearson-Bravais zwischen vorhergesagten und beobachteten Werten beträgt 0,298 und ist einseitig hochsignifikant.

5.6.4.3 Ergebnisse des Regressionsbaumes

Für die durch den Regressionsbaum vorhergesagten Werte ist $R_1 = 0,294$, dies entspricht 97 richtig vorhergesagten von insgesamt 330 Werten.

Ähnlich wie der Klassifizierungsbaum sagt der Regressionsbaum etwa 2/3 der Entscheidungen der Versuchspersonen hinsichtlich der Identität der Testmelodie richtig voraus: $R_2 = 0,664$, das entspricht 219 richtigen von 330 Beobachtungen insgesamt.

Dieser Zusammenhang wurde durch eine χ^2 -Test als hochsignifikant ausgewiesen ($\chi^2 = 37,162$, $df = 1$, $p < 0,001$).

Für die mittlere quadrierte Abweichung ergibt sich ein Wert von $R_3 = 0,945$. Gemessen an diesem Kriterium sind die Vorhersagen des Regressionsbaumes also leicht besser als die einfache Verwendung des Mittelwertes. Der korrelative Zusammenhang zwischen beobachteten und durch den Regressionsbaum vorhergesagten Werten ist 0,358 und hochsignifikant.

5.6.4.4 Ergebnisse des zweistufigen Klassifizierungsbaumes

Auf dem Maß R_1 erreicht der zweistufige Klassifizierungsbaum einen Wert von 0,318, was 105 von 330 richtig klassifizierten Beobachtungen entspricht.

Auch bei der Vorhersage der dichotomen Urteile identisch/verschieden reiht sich das zweistufige Modelle – oder genauer: dessen erste Stufe – in den Bereich der richtigen Vorhersagen von etwa 2/3 ein: 65,5% der Beobachtungen werden richtig vorhergesagt, dies entspricht einem R_2 -Wert von 0,655. Der χ^2 -Test erbrachte einen Wert von 32,686, was bei einem Freiheitsgrad einen hochsignifikanten Zusammenhang darstellt ($p < 0,001$).

Der Wert für R_3 ist dagegen deutlich schlechter als bei den Konkurrenzmodellen. Er beträgt 1,092. Die bivariate Korrelation zwischen den durch dieses Modell vorhergesagten und den beobachteten Werten ist mit 0,242 denn auch deutlich geringer, obwohl immer noch hoch signifikant.

Die folgende Tabelle fasst die Ergebnisse der drei Modelle hinsichtlich der verwendeten Maßzahlen zusammen.

Maß	Lineares Modell	Klassifizierungsbaum	Regressionsbaum	2-stufiger Regressionsbaum	Häufigste Kategorie bzw. Mittelwert
R	0,321	0,321	0,294	0,318	0,248
R₂	0,645	0,655	0,664	0,655	0,518
R₃	1,002	1,062	0,945	1,092	1

Tabelle 5-25: Zusammenfassung der Evaluationsergebnisse der vier verschiedenen Vorhersagemodelle.

5.6.5 Diskussion der statistischen Modellierung

Wie die vorangegangene Tabelle zeigt, ist die Vorhersagequalität keines der drei Modelle überragend. Alle Modelle erzielten in etwa die selben Vorhersageergebnisse und bringen grob gesagt doppelt so gute Ergebnisse wie das Raten, was sich auch in den oben referierten deutlichen Signifikanzwerten der Zusammenhänge mit den beobachteten Werten ausdrückt.

Je nachdem, wie in diesem Fall Prognosequalität definiert wird, ist entweder der Klassifizierungsbaum oder der Regressionsbaum zu wählen. Geht es beispielsweise darum, in einer Wette auf die richtige Vorhersage des *rating*-Wertes zu wetten, so sollte der Klassifizierungsbaum oder das lineare Modell gewählt werden, die in 32% der Fälle richtig liegen.

Dies ist doppelt so gut, wie die einfache Ratewahrscheinlichkeit von $1/6$. Ist die Vorhersage der Entscheidung identisch/verschieden gefragt, würde man den Regressionsbaum wählen, der ziemlich genau in $2/3$ der Fälle richtig liegt. Eine faire Wette, würde einen Einsatz von 2:1 anbieten gegenüber einem Modell, das die Entscheidung der Vpn. ohne Vorinformation rät. Auch dieses Modell kann man als doppelt so gut wie die einfache Ratewahrscheinlichkeit bezeichnen. Das lineare Modell ist in keinem Fall einem der beiden Baummodelle überlegen.

Bei einer Wette, bei der die Abweichungen der gewetteten (vorhergesagten) von den beobachteten Werte Verluste bedeuten, würde man in jedem Fall den Regressionsbaum verwenden, der den deutlich besten Wert für R_3 aufweist.

Das zweistufige Baummodell, bei welchem zuerst die dichotomen Urteile und danach die Urteilssicherheit vorhergesagt werden, hat sich in keinem der Fälle als effektiv erwiesen. Zum einen kann das Modell zur Urteilssicherheit nicht die Vorhersagewerte zu den dichotomen Urteilen verwenden, wie die grafischen Darstellungen oben zeigen. Zum anderen weist es bei keinem der drei Modellgütekriterien einen besonders guten Wert auf. Es macht offenbar bei den hier gegebenen Prädiktoren keinen Sinn, die Urteilssicherheit unabhängig vom Urteil vorherzusagen.

Diese Ergebnisse zeigen zum einen, dass sich die Modellierung durch die *data-mining*-Ansätze gelohnt hat. Eine Vorhersage der Entscheidung der Versuchspersonen, ob eine Testmelodie identisch oder verschieden ist und des *rating*-Wertes der Urteilssicherheit ist möglich und bringt doppelt so gute Ergebnisse wie das einfache Raten. Diese Vorhersage stützt sich dabei nicht auf Vorwissen um die Art und die Stärke der Veränderung der Testmelodie (*Manipulationsklasse*, *Manipulationstyp* sowie *Akzentstärke* der beteiligten Töne), die durch den Experimentator festgelegt wurden, sondern die Größen können gemessen und als Prädiktor verwendet werden. Dies ist ein großer epistemologischer Vorteil gegenüber dem vierfaktoriellen Modell, welches in 5.5.2 aufgrund der Konstruktionsprinzipien der Versuchsmaterialien aufgestellt wurde. So können beispielsweise Materialien und Daten aus ähnlichen Situationen oder wissenschaftlichen Experimenten anderer Forscher ohne Vorwissen um deren Konstruktion analysiert und die Reaktion der Versuchspersonen darauf vorhergesagt werden. Beispielsweise wäre eine Analyse der aus dem Gedächtnis wiedergegebenen Melodien der Versuchspersonen von Sloboda und Parker (1985) möglich, die dort in Notentranskription veröffentlicht sind. Auch in Hinsicht auf die Modellsituation einer *Query-by-humming*-Anwendung, die ja Leitvorstellung dieser Arbeit ist,

ist die Möglichkeit zur Vorhersage einer menschlichen Reaktion ohne Kenntnis der Prinzipien, die zur Konstruktion einer konkreten Melodie geführt haben, sinnvoll. Es müssen nur die Referenzmelodie(n) und die aktuelle Testmelodie (*query*) bekannt sein. Die algorithmischen Maße *opti1* und *joint52* liefern dann für die Vorhersage verwendbare Daten durch Messung der melodischen bzw. der Akzentähnlichkeit.

Zum anderen weisen die Werte für die Modellgüte der drei Modelle daraufhin, dass a) noch eine Reihe weiterer Faktoren das Urteil der Versuchspersonen beeinflussen oder b) die Entscheidung der Vpn., welcher Wert auf der Urteilsskala anzukreuzen ist von einem wie auch immer zu definierenden Zufall abhängt oder c) die einzelnen Versuchspersonen nach verschiedenen Kriterien urteilen bzw. die Kriterien verschieden gewichten. Es lässt sich anhand der erhobenen Daten nicht entscheiden, welche Ursache tatsächlich verantwortlich für die Güte der vorliegenden Modelle ist. Möglich ist allerdings, die Modelle durch Hinzunahme von Prädiktoren zu verbessern, die beispielsweise andere, zusätzliche inhaltliche Informationen liefern.

Sinnvoll könnte hier ein kombiniertes aus melodischer Ähnlichkeit und Akzentähnlichkeit kombiniertes Maß sein. Dieses Maß würde wahrscheinlich eine Gewichtung der (Un-)Ähnlichkeiten zweier Melodien anhand der Übereinstimmungen ihrer Akzenttöne vornehmen. Ein Optimierungsprozess auf einem geeigneten Datensatz, der nicht zu diesem Experiment gehört, mit den aussichtsreichsten Kandidaten zur Messung sowohl melodischer als auch Akzentähnlichkeit, wäre hier von Nöten.

Ein weiterer wichtiger Faktor könnte die Seltenheit oder Gängigkeit (vgl. Steinbeck, 1982, S. 230ff) der Testmelodie sein. Es hatten sich ja schon in der ersten Auswertung starke Unterschiede zwischen in der Beurteilung der Testmelodien als identische oder verschiedenen unabhängig von ihren Versuchsbedingungen gezeigt (vgl. Abschnitt 5.4.7). Trotz der Berücksichtigung bestimmter Eigenheiten der Testmelodien, wie ihrer Länge oder der Ähnlichkeit in der Akzentstruktur zur Referenzmelodie, sind die spezifischen Unterschiede zwischen den Testmelodien immer noch feststellbar. Betrachtet man die Residuen des Regressionsbaummodells auf dem Testdatensatz, so zeigen sich einige Auffälligkeiten, wie die *boxplots* der Verteilung der Residuen zu den einzelnen Melodien zeigen. Es sind hier nur die Residuen der Melodien 1 bis 6 dargestellt. Die Residuen der übrigen Melodien finden sich im Anhang.

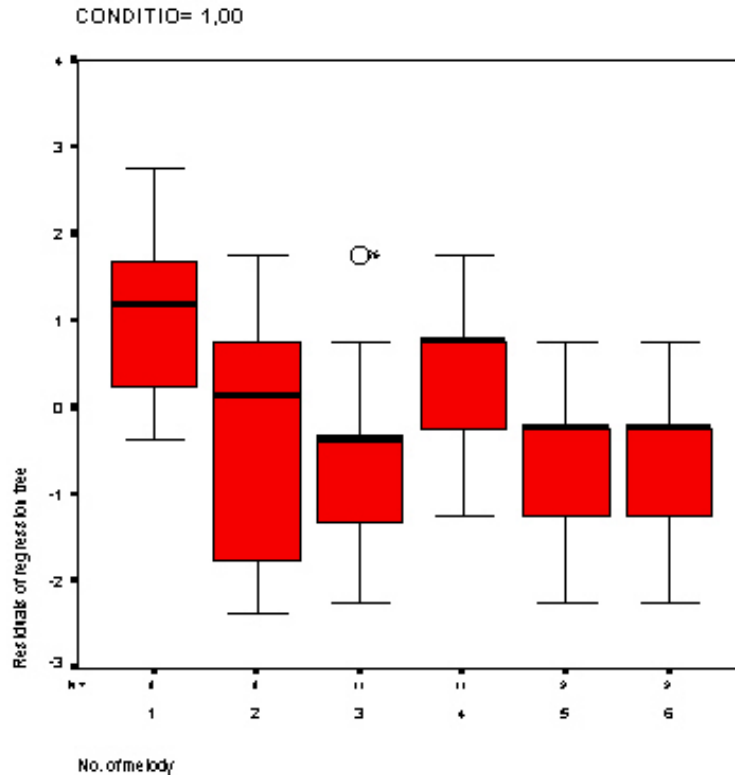


Abb. 5-12: Residuen des Regressionsbaummodells zu den Melodien 1 bis 6. Die schwarzen Balken repräsentieren den Median der Verteilung, das große Quadrat (*box*) beschreibt die Fläche vom 25%-Quantil bis zum 75%-Quantil. Die waagerechten dünnen Linien geben den Wert des ersten bzw. letzten Falles an, der kein Ausreißer ist, d.h. um mindestens das 1,5fache der Boxhöhe vom 25%- bzw. 75%-Perzentil entfernt ist. Der Kreis bezeichnet eine Beobachtung, die einen Ausreißer darstellt.

Wären die nicht erklärten Anteile an den Urteilen der Versuchspersonen nur statistisches ‚Rauschen‘ wären, müssten die Residuen annähernd normalverteilt um den Median streuen, der in der Nähe des Nullpunktes liegen müsste. Wie jedoch schon an dieser Abbildung deutlich zu sehen ist, sind die Mediane bei einigen Melodien deutlich von 0 verschieden (z.B.: Median zu Melodie 1: 1,1895; Melodie 10: -1,147, Melodie 18: 1,55, Melodie 28: 1,853; s. Anhang), und die Verteilung der Residuen streut nur in den wenigsten Fällen annähernd normal um dem Median herum. Um zu prüfen, ob die Modellresiduen aus der selben oder aus unterschiedlichen Verteilungen stammen, stehen nicht-parametrische Tests zur Verfügung. Der Kruskal-Wallis-Test prüft die Zusammenhänge zwischen den Verteilungen für unabhängige Stichproben, der Friedmann-Test tut dies für abhängige Stichproben. Bei der hier gegebenen Datenlage ist nicht eindeutig entscheidbar, ob die Residuen unabhängige oder abhängig voneinander sind. Denn in die Residuen zu zwei zu vergleichenden Melodien können sowohl die Daten der selben Vpn. eingehen als auch die Daten von verschiedenen Versuchspersonen. Dies ist dadurch begründet, dass nicht alle Versuchspersonen mit allen Melodien getestet wurden, sondern nur mit zwei Dritteln der gesamten Testmelodien. Aus diesem Grund wurden sowohl der Kruskal-Wallis- als auch der

Friedmann-Test durchgeführt. In beiden Tests wurden die Verteilungen der Residuen zu allen 36 Melodien gegeneinander getestet. Beide Tests ergaben signifikante Werte bei zweiseitiger Prüfung (Kruskal-Wallis: $p = 0,009$; Friedmann: $p = 0,012$), d.h. die Residuen stammen (unter Annahme der angegebenen Irrtumswahrscheinlichkeiten) aus unterschiedlichen Verteilungen. Dies wiederum kann inhaltlich dahingehend interpretiert werden, dass die einzelnen Testmelodien oder die getesteten Musikstücke noch Eigenschaften enthalten, die nicht durch die hier als Prädiktoren verwendeten Messgrößen erfasst wurden. In jedem Fall enthalten die nicht-erklärten Anteile der Versuchspersonenurteile nicht ausschließlich Rauschen, sondern mit den Eigenschaften der Melodien oder Teststücke verbundene Information, die prinzipiell zur Optimierung der statistischen Vorhersagemodelle zur Verfügung steht.

Als Vermutung drängen sich hier die spezifischen Eigenschaften der Melodien selbst auf, und zwar im Sinne von deren ‚Plausibilität‘. Angenommen wird dabei, dass sich die Vpn. in der Testsituation folgende Frage stellt: „Kann es so eine Testmelodie in einem Musikstück tatsächlich geben oder besitzt die Testmelodie so merkwürdige Eigenschaften, dass es eher wahrscheinlich ist, dass sie eine Manipulation einer anderen, geläufigeren Melodie ist?“ Es scheint möglich die Versuchspersonen in dieser experimentellen Urteilsituation mit einem Bayesianischen Ansatz zu modellieren. Neben der gedanklich mit Sicherheit anspruchsvollen Ausarbeitung dieses Ansatzes, wäre dafür zudem eine große Datenbank mit Popmusikmelodien notwendig, aus der sich typische melodische Muster ableiten ließen.

Zusätzlich ist auch zu bedenken, dass die Schwierigkeiten der Vorhersage auf der im Wertebereich begrenzten Ordinalskala – trotz des Versuches diese mit einem zweistufigen Baummodell zu meistern – systematisch die Vorhersagequalität beeinträchtigen: Der Klassifizierungsbaum kann die Ordnungsinformation bei der Vorhersage der *rating*-Werte nicht ausnutzen.

Das auf dem kleinste-Quadrate-Kriterium beruhende Baummodell (Regressionsbaum) hat dagegen Schwierigkeiten durch die Begrenzung der Skala. Da Gruppenmittelwerte vorhergesagt werden, können nur sehr schwer die Werte an den Skalenenden vorhergesagt werden. Der Vorhersagebereich ist in diesem Fall auf den Bereich zwischen 2,45 und 4,375 beschränkt. Hier böte sich an, den in einem Endknoten zusammengefassten Beobachtungen nicht den Knotenmittelwert als Vorhersagewert zuzuordnen, sondern zum Beispiel den Modus des Knotens. Da diese Möglichkeit aber nicht in der hier verwendeten Software

vorgesehen ist, muss die Frage, um wie viel sich die Vorhersagequalität des Regressionsbaums durch eine solche alternative Benennung der Endknoten verbessert hätte, erst einmal unbeantwortet bleiben.

Eine Re-Analyse der Daten mit einem Modell, welches die Eigenheiten der hier verwendeten abhängigen Variablen berücksichtigt, wäre jedoch in jedem Fall sinnvoll.

Abgesehen von der Tatsache, dass die Modelle eine gewisse Vorhersagequalität besitzen und in analogen Situationen zur Vorhersage eingesetzt werden können, ist es durch die Interpretation der Modelle auch möglich, etwas über die kognitiven Mechanismen zu erfahren, die bei der Bewältigung der Testaufgabe eine Rolle gespielt haben. Da die beiden Baummodelle sich hinsichtlich der Gütekriterien kaum unterscheiden bzw. sogar leicht besser sind und zudem den Vorteil besitzen, dass sie leichter interpretierbar sind als das lineare Modell mit seinen zahlreichen Interaktionstermen, soll die Interpretation zunächst von den Baummodellen ausgehen. Da diese beiden Modelle zusätzlich inhaltlich noch relativ ähnlich sind, ist eine zusammenfassende Interpretation beider Modelle möglich.

Das größte *improvement* im Sinn der Verbesserung des Unreinheitsindex (Gini) bzw. der erklärten Varianz brachte in beiden Modellen die Teilung in identisch und nicht-identische Melodien. In beiden Modellen ist die durch in Kapitel 4 beschriebene optimale Ähnlichkeitsmaß *opt1* die Variable, die den Wurzelknoten teilt. Im linken Ast des Regressionsbaums werden alle nicht-identischen Melodien weiterhin ausschließlich durch diese Variable und durch die inhaltlich verwandte Akzentähnlichkeit *joint52* unterteilt. Die objektiven Übereinstimmungen zwischen Testmelodie und Referenzmelodie wie sie durch *opt1* und *joint52* messbar sind, erweisen sich also als wichtigster Prädiktor.

Zwei weitere Gruppen von Variablen sind in den Baummodellen bedeutsam. Subjektbezogene Variablen, die den musikalischen Hintergrund der Versuchspersonen beschreiben und Variablen, die durch das Design des Experimentes festgelegt werden sind und angeben, wie lang die Testmelodie ist bzw. welchen Abstand sie von der Referenzmelodie hat. Im Fall der subjektbezogenen Variablen zeigt sich der erwartete positive Zusammenhang zwischen der musikalischen Aktivität der Versuchspersonen und einer höheren und häufiger korrekten Einschätzung der Identität der Testmelodien im Regressionsbaum: Je musikalisch aktiver die Versuchspersonen sind, desto besser schätzen sie die identischen Melodien ein. Im Gegensatz dazu, führt nach dem Klassifizierungsbaum ein mittlerer Musik-

konsum zu den besten Ergebnissen: Wer sehr wenig Musik oder sehr viel Musik hört, urteilt mit geringerer Urteilssicherheit, zumindest was die nicht-identischen Melodien angeht.

Die Richtung des Zusammenhangs zwischen Länge bzw. Dauer der Testmelodie und der Urteilssicherheit ist dagegen recht eindeutig. Identische Melodien werden dann eher und eindeutiger als solche beurteilt, wenn sie länger sind: *split* 6 des Regressionsbaumes mit Testmelodien von weniger als 16,5 Tönen besitzt einen Vorhersagewert von 3,49, und *split* 7 mit längeren Testmelodien zeigt 4,24 als Vorhersagewert an (vgl. die entsprechenden Tabellen im Anhang). Ganz analog verhält sich mit der Dauer der Testmelodien, die Prädiktor im Klassifizierungsbaum ist. Dieser positive Zusammenhang zwischen Ausdehnung der Testmelodie und der Tendenz, ihre Identität sicherer einzuschätzen, lässt sich plausibel dadurch erklären, dass in längeren Melodien den Versuchspersonen mehr Material zum Wiedererkennen dargeboten wird bzw. sie mehr Zeit dafür haben, dass sich das Gefühl einstellt, etwas als identisch wiedererkannt zu haben.

Wagt man trotz der größeren Komplexität einen Blick auf das lineare Modell, so stellt man fest, dass auch dort diese drei Gruppen von Variablen – melodische und Akzentähnlichkeit, musikalischer Hintergrund der Versuchspersonen und Ausdehnung der Testmelodie – in den allermeisten Interaktionstermen als Faktor vorhanden sind und in der Regel auch die selbe Richtung des Zusammenhangs aufweisen, wie sie in den Baummodellen gefunden wurde. Auch wenn man das lineare Modell keiner detaillierten analytischen Interpretation unterwirft, so stützen doch auch die in ihm verwendeten Variablen auf den ersten Blick, die an den beiden Baummodellen gewonnenen Interpretationen und Erkenntnisse.

Die Tatsache, dass die messbare melodische Ähnlichkeit der beste Prädiktor ist, lässt sich als positives Ergebnis werten, denn dieser lässt sich in vielen weiteren experimentellen und anwendungsorientierten Situationen einsetzen, während viele andere Größen in Anwendungssituationen oft unbekannt sind. So ist z.B. in einer Internetanwendung eines *Query-by-humming*-Systems in der Regel nichts über die musikalische Aktivität und den musikalischen Hintergrund der Benutzer bekannt. Genauso wenig dürfte in vielen Situationen darüber bekannt sein, wie viel Zeit zwischen dem Hören einer Melodie und der Reproduktion bzw. dem Wiedererkennensvorgang vergangen ist oder wie viele Melodien oder Töne der Benutzer in der Zwischenzeit gehört hat.

Ein großer Vorteil der Baummodelle besteht auch darin, dass durch Ermittlung der Surrogat-Variablen bzw. der nächst besten Variablen für einen *split* in einem bestimmten Kno-

ten, bei Unbekanntheit einer Variablen relativ einfach und ohne zusätzliche Berechnung ein alternatives Modell verwendet werden kann, was ohne die unbekannten Variablen auskommt.

Beispiel ist eine *Query-by-humming*-Anwendung, in der nur Variablen verfügbar sind, die die Unterschiede zwischen Referenz- und Testmelodie oder die Ausdehnung der Testmelodie beschreiben, also *similar*, *joint52*, *harmfunc*, *changers*, *durtemel*, *lentemel*, *noterest*, *manrhyth*, *maninter*, *mancont*. Die Knoten 1 (Wurzel), 2 und 3 des Regressionsbaumes bleiben unverändert, da sie *similar*, *joint52* und *lentemel* verwenden. Im vierten Knoten wäre die Regel ‚*zmusicac* < 0,312‘ durch die nächst schlechtere ‚*durtemel* < 9,5‘ zu ersetzen. Dies halbiert zwar die Verbesserung des ANOVA-Kriteriums in diesem Knoten von 0,135 auf 0,07, macht aber ein vollständiges Baummodell für diese spezielle Analysesituation verfügbar.

Der Sinn dieser ‚Übung‘, d.h. die in den vorangegangenen Abschnitten beschriebene Modellierung mit Ansätzen des *data mining*, diene nicht nur der bestmöglichen Vorhersage der *rating*-Daten dieses Experimentes und der Evaluierung der Bedeutung der einzelnen Prädiktorvariablen, sondern auch der Evaluation von Methoden zur Modellierung von experimentellen Daten aus musikpsychologischen Experimenten zum Musikgedächtnis ganz allgemein.

Die meisten Experimente, die Gedächtnisleistungen für ‚echte‘, nicht-künstliche musikalische Materialien untersuchen, generieren Datensätze, die durch eine Reihe von typischen Eigenschaften zu charakterisieren sind. Diese typischen Eigenschaften, die zwar nicht vollständig bei allen Experimenten dieser Art vorliegen, aber doch bei der überwiegenden Mehrheit vorhanden sind, lassen sich stichpunktartig wie folgt zusammenfassen:

- Inhaltlich ist eine Beteiligung von vielen Prädiktorvariablen (multivariater Ansatz, hohe Dimensionalität) auf allen Skalenniveaus plausibel.
- Die Prädiktorvariablen besitzen häufig fehlende Werte.
- Wechselwirkungen auch höheren Grades zwischen den Prädiktoren sind inhaltlich denkbar und i.d.R. nicht auszuschließen.
- Nicht-lineare Zusammenhänge zwischen der abhängigen und einer oder mehreren unabhängigen Variablen sind oft nicht auszuschließen.
- Oft sind die Voraussetzungen hinsichtlich der Verteilungen von abhängiger und unabhängiger Variable nicht gegeben, die bestimmten Klassen von Modellen for-

dern (z.B. Normalverteilung der Residuen oder verschiedene große Fallzahlen für verschiedene Ausprägungen einer Variable).

- Nicht durch das Experimentaldesign kontrollierbare, aber bedeutsame Einflüsse, die durch die Verwendung echten musikalischen Materials zustande kommen, müssen berücksichtigt werden.
- Ein sehr heterogener kognitiver Hintergrund der Versuchspersonen für die Bewältigung der Testaufgabe ist häufig gegeben. Der kognitive Hintergrund kann – wie in diesem Beispiel – für die Wahl von verschiedenen Bewältigungsstrategien verantwortlich sein.

Durch die Konstruktion sowohl eines linearen wie auch zweier bzw. dreier Baummodelle, die in der vorangegangenen Abschnitten erläutert wurden, sind die Schwierigkeiten und die Lösungsmöglichkeiten aufgezeigt worden, die sich bei Verwendung der beiden unterschiedlichen Modellansätze mit dieser Art von Daten ergeben. Wie die Maße zur Modellgüte ergeben haben, sind die beiden Ansätze in ihrer Vorhersagequalität für das vorliegende Problem vergleichbar, mit einem leichten Vorteil für die Baummodelle. Jedoch zeigt sich, dass die Baummodelle deutlich weniger ‚Konstruktionsarbeit‘ verursachen und ihre Interpretation und Handhabung bedeutend einfacher ist als die des linearen Modells. Dies ist begründet durch die Mechanismen der Baummodelle, die mit den aufgezählten Eigenschaften bzw. Schwierigkeiten der generierten Datensätze vergleichsweise einfach umgehen.

So stellt eine hohe Anzahl von Prädiktoren und die Abbildung von Wechselwirkungen auch hoher Ordnung sowie von nicht-linearen Zusammenhängen kein Problem für das Prinzip *Rekursive Partitionierung des Antworttraumes* dar, welches den Baummodellen zugrunde liegt. Abhängig von der Wahl der Parameter, z.B. des Komplexitätsminimums, können die Baummodelle sehr sparsam in der Verwendung von Prädiktoren und *split*-Regeln sein. Dies erleichtert die Interpretation der Modelle ungemein. Im Gegensatz dazu enthielt das lineare Modell Terme mit Vorzeichen, die so nicht erwartet worden und dadurch schwer zu interpretieren waren. Die Interaktion von Variablen wird im Baummodell durch die Verknüpfung der in der rekursiven Partitionierung verwendeten Variablen abgebildet. Dies ist in vielen Situationen intuitiv nicht schwer zu begreifen und zu vermitteln. Im Gegensatz dazu ist die Multiplikation als Verknüpfung zweier Variablen in einer Interaktion nur eine, wenn auch die geläufigste Möglichkeit, die jedoch den Interpretierenden oft vor Schwierigkeiten stellt. Die Ordnung der Interaktionen wurde bei der Konstruktion des linearen Modells auf die zweite Ordnung beschränkt. Bei der Betrachtung von Interak-

tionstermen höheren Grades und der Verwendung von vielen Prädiktoren wächst die Anzahl der möglichen Terme schnell in Bereiche, die in akzeptabler Zeit nicht mehr zu berechnen sind. So existieren für k Variablen 2^k Teilmengen, d.h. mögliche Interaktionsterme, wenn keine Beschränkung auf eine Ordnung der Interaktionen festgelegt ist. Bei 19 Prädiktoren enthielte ein vollständiges Modell mit Hauptfaktoren und Interaktionstermen 1. Ordnung 190 Modellterme, zuzüglich der Interaktionsterme 2. Ordnung wären es schon 1159 Terme insgesamt. Der Zusammenhang zwischen der Ordnung der berücksichtigten Interaktionsterme und den möglichen Modelltermen ist exponentiell, d.h. für Interaktionsterme 3. Ordnung ist er kubisch, für solche 4. Ordnung quadratisch etc. Die Variablenselektion wird dadurch extrem aufwendig oder es wird gar unmöglich, ein vollständiges Modell zu notieren, geschweige denn das richtige Modell mit Routinen zur Variablenselektion zu identifizieren.

Die Baummodelle sind relativ voraussetzungsarm, so sind beispielsweise für die Festlegung von split-Kriterien in einem Regressionsbaum die ordinalen Informationen der Prädiktoren ausreichend. Annahmen hinsichtlich der Verteilung von vorhergesagten Werten oder Residuen werden nicht gemacht. Die gleichzeitige Berücksichtigung von intervallskalierten, ordinalen und kategorialen Prädiktoren ist unproblematisch. In linearen Modellen sind hierfür oft weniger transparente Umkodierungen in Dummy-Variablen notwendig. Schließlich dürfte der größte Vorteil der Baummodelle ihr Umgang mit fehlenden Werten in den Prädiktorvariablen durch die Verwendung von Surrogat-Variablen sein. Es ist hier also keine Entscheidung von Nöten, ob fehlende Werte durch ein geeignetes Verfahren ersetzt werden sollen (z.B. Mittelwert des Prädiktors oder Mittelwert der irgendwie benachbarten Beobachtungen) oder ob auf die Vorhersage der Beobachtung in einem solchen Fall verzichtet werden muss.

Gerade auch aus diesen praktischen und interpretatorischen Erwägungen ist das Baummodell hier als Erklärungsmodell gewählt worden. Umso deutlicher die Vorteile des Baummodells in diesem Fall auf der Hand liegen, desto erstaunlicher ist es, dass sie bisher bei der Analyse von Daten in musikpsychologischen Experimenten Baummodelle keinerlei Anwendung gefunden haben. Eine Spekulation über die Gründe für die wenig verbreitete Verwendung von Baummodellen soll hier dennoch unterbleiben.

5.7 Zusammenfassung und Diskussion

Das in diesem Kapitel beschriebene Experiment sollte Erkenntnisse zu einer Reihe von verschiedenen Bereichen erbringen. Die wichtigsten Felder sollen zum Abschluss noch einmal angesprochen und die zugehörigen Erkenntnisse aus diesem empirischen Kapitel zusammengefasst werden.

Verwendung von erprobten experimentellem Design unter veränderten Voraussetzungen:

Ein bekanntes experimentelles Design von Wiedererkennensexperimenten zum Musik- und Melodiegedächtnis sollte bei Verwendung von populären echten Musikstücken auf seine Tragfähigkeit überprüft werden. Konkret standen die Experimentaldesigns von Kauffmann und Carlsen (1989) sowie Dowling et al. (2002) Pate bei der Wahl des Designs und der experimentellen Parameter. Das vorliegende Experiment geht über diese beiden Studien jedoch in vielerlei Hinsicht hinaus: So war das Musikmaterial den allermeisten Versuchspersonen stilistisch sehr vertraut, da es sich um Pop-, Rock- und Schlagertitel handelte, die natürlich melodische und harmonische Wendungen und Konstruktionsprinzipien enthielten, die den Versuchspersonen zumindest implizit sehr geläufig waren. Im Gegensatz dazu verwendeten Kauffman und Carlsen sowie Dowling und Kollegen Ausschnitte aus klassischen und romantischen Kompositionen, deren Konstruktionsprinzipien den meisten ihrer Versuchspersonen weniger geläufig gewesen sein dürften.

Es wurde damit bewusst in Kauf genommen, dass die Versuchspersonen Assoziationen oder erlerntes Wissen über typische musikalische Verläufe in diesem Stil bei der Bewältigung der Versuchsaufgabe einfließen lassen. Maßgabe war hier die Speicherung von populären Melodien im Gedächtnis, die der Reproduktion in *Query-by-humming*-Situationen vorausgeht. Auch in der *Query-by-humming*-Situation ist nicht davon auszugehen, dass ein Benutzer sich eine Melodie oder einen Melodieausschnitt ohne Kontext, quasi in einem nicht vorgeprägten Gedächtnis, merkt. Vielmehr dürften auch hier typische musikalische Vorerfahrungen eine Rolle spielen.

Zusätzlich zur Verwendung von musikalischem Material, das den Versuchspersonen besser bekannt ist, war im Gegensatz zu beiden vorgenannten Studien von den Versuchspersonen eine Abstraktion der Melodie aus dem Musikstück gefordert. Die Testmelodie in der Wiedererkennensphase wurde hier jeweils einstimmig dargeboten, während in den beiden früheren Studien ein vollständiger Ausschnitt aus dem Teststück inklusive Begleitung zu hören war. Diese Forderung nach Abstraktion der Melodien war in Hinsicht auf die Über-

tragbarkeit der Ergebnisse für *Qbh*-Situationen notwendig. Anhand der gewonnenen Daten ist deshalb nicht zu entscheiden, inwieweit die gemessenen Gedächtnisleistungen bzw. das Antwortverhalten der Versuchspersonen auch z.T. auf deren individuelle Fähigkeit zur Melodieabstraktion zurückgehen. Dies ist für die Übertragung der Ergebnisse für mögliche *Qbh*-Anwendungen zwar auch nicht von Bedeutung, könnte jedoch eine sehr interessante Fragestellung für eine Nachfolgeuntersuchung sein.

Insgesamt lässt sich die Anwendung des erprobten experimentellen Designs mit den bekannteren Stimuli und der zusätzlichen Anforderung der Melodieabstraktion als erfolgreich bezeichnen. Es ließ sich ein Modell zur Vorhersage aufstellen, welches signifikant besser vorhersagt als das Werfen eines Würfels und auch zumindest unter manchen Aspekten deutlich besser ist als die Verwendung von konstanten Maßzahlen aus früheren Versuchen des selben Typs. Zusätzlich kommt das Modell auch unter den bewusst veränderten Bedingungen zu ähnlichen Ergebnissen, wie sie auch Dowling et al. (2002) und Kauffman und Carlsen (1989) fanden. Es besteht in dieser Hinsicht also kein Grund an der Plausibilität der Ergebnisse zu zweifeln.

Wahl der richtigen Antwortgröße (*response variable* = abhängige Variable):

Bei der Modellbildung anhand der kontrollierten experimentellen Parameter zeigte sich, dass – im Gegensatz zum üblichen Vorgehen wie beispielweise bei Dowling et al. (2002) – sich nicht die Gedächtnisleistungen der Versuchspersonen als die sinnvolle Antwortgröße erwiesen, sondern das Kreuzverhalten auf der sechsstufigen Skala. Bei der Verwendung der Gedächtnisleistungen, die über die Maße AUC und d' aus der Signal Detection Theory operationalisiert werden, konnte kein multivariates Modell erstellt werden, dass auch nur annähernd zufriedenstellende Indikatoren der Modellgüte oder signifikante Prädiktoren aufwies ($R^2 = 0,058$, bzw. $R^2 = 0,078$). Dagegen zeigte ein Modell mit den selben Prädiktoren und dem ‚Kreuz auf der sechsstufigen Antwortskala‘ als abhängige Variable akzeptable Modellwerte an ($R^2 = 0,199$).

Es zeigt sich also bei der Verwendung von bekannterem musikalischen Material und einer insgesamt schwierigeren Versuchsaufgabe, dass es in erster Linie sinnvoll ist, die Antworten der Versuchspersonen vorherzusagen und so ihr Antwortverhalten zu verstehen. Auch aus epistemologischen Gründen kann es in diesem Fall sinnvoll sein, die Antwortgröße nicht von der Entscheidung des Experimentators abhängig zu machen, was bei einem gegebenen *item* als richtig und falsch zu bewerten ist. Denn möglich wäre, dass Experimentator und Vpn. in ihrer Bewertung, was identisch oder verschieden im vorliegenden Ver-

such bedeutet, auch unter idealen Bedingungen nicht übereinstimmen. Für eine Vpn. mag eine minimale rhythmische Abweichung nichts an der Identität zweier Melodien ändern, für den Versuchsleiter ist die Abweichung aber vielleicht der Grund, die beiden Melodien als verschieden zu werten. Bei der Betrachtung der Kreuze auf der sechsstufigen Skala erübrigt sich eine vorherige Festlegung der Kategorien ‚richtig‘ und ‚falsch‘.

Trotzdem können innerhalb eines solchen Modells im nachhinein immer noch die Entscheidungen der Versuchspersonen analysiert werden, die aus Sicht des Forschers zu falschen Entscheidungen der Versuchspersonen geführt haben. Dies wäre unter Annahme des Regressionsbaumes der Endknoten mit dem Mittelwert 3,246, der sich bei identischen Testmelodien (*similar* $\geq 0,8603$) ergibt: ‚Fehlentscheidungen‘ lassen sich voraussagen für Versuchspersonen mit geringer musikalischer Aktivität bei identischen Testmelodien, die kurz sind, d.h. weniger Noten enthalten.

Identifizierung bedeutsamer Einflussgrößen für das Wiedererkennensgedächtnis von Melodien:

Die üblicherweise zentrale Fragestellung musikpsychologischer Experimente ist die nach der Wirkungsweise verschiedener musik- oder subjektbezogener Größen auf die Ergebnisse eines Experimentes. Beim vorliegenden Experiment bietet sich eine Interpretation des einfachsten und trotzdem leistungsfähigen Regressionsbaummodells an, wie sie weiter oben geleistet wurde. An erster Stelle beeinflussen die melodische Ähnlichkeit und die Übereinstimmung in der Akzentstruktur das Urteil der Versuchspersonen. Nach diesen Übereinstimmungen lassen sich die Antwortmittelwerte aller nicht-identischen Melodien innerhalb des Modells vorhersagen. Nach der Akzentähnlichkeit ist die Länge der Testmelodie entscheidend: Je länger sie ist – und wahrscheinlich je mehr Material sie zum Vergleich mit der internen Gedächtnisrepräsentation liefert – desto eher wird ihre Identität erkannt. An dritter Stelle steht schließlich die musikalische Aktivität der Versuchspersonen in der Vergangenheit. Ist diese geringer, neigen die Versuchspersonen eher dazu, identische Melodien als unterschiedlich zu bewerten. Möglicherweise ist dies durch eine geringere Kapazität des Melodiegedächtnisses bei diesen Versuchspersonen zu erklären, die große Teile der Referenzmelodie bereits vergessen haben und so deren Identität nicht bestätigen können.

Über die in dem Vorhersagemodell verwendeten Prädiktoren hinaus geben einige der gefundenen Ergebnisse auch Anhaltspunkte für bisher unberücksichtigte Einflussfaktoren, deren Berücksichtigung noch zu einer Verbesserung der Modellgüte beitragen kann. Sehr

deutliche Unterschiede treten bei der Bewertung von verschiedenen Melodien unter an sich gleichen experimentellen Bedingungen und auch nach der Berücksichtigung einiger Merkmale der melodischen Struktur im statistischen Modell zu Tage. Dies ergab die Untersuchung der Residuen des Regressionsbaummodells, die offensichtlich für die einzelnen Melodien aus unterschiedlichen Verteilungen stammen. Diese Unterschiede können nur durch weitere Prädiktoren erklärt werden, die die charakteristischen Eigenheiten der Melodien in Hinsicht auf das vorliegende Experimentaldesign adäquat abbilden.

Als Kandidaten kommen hier vor allem die schon erwähnten möglichen Informationen über die Gängigkeit oder Sinnfälligkeit der Testmelodien in Betracht. Dass diese Größen eine bedeutende Rolle spielen können, zeigen auch die von den Versuchspersonen eingesetzten Gedächtnisstrategien, die in den kurzen qualitativen Befragungen berichtet wurden. Hier gingen einige Strategien vom Vorwissen über die Konstruktion von Melodien und Rhythmen aus. Es wurden sowohl prägnante Motive mit vorbekanntem Material assoziiert als auch die Testmelodien aufgrund ihrer Sinnfälligkeit, d.h. ihrer Plausibilität im Vergleich mit den melodischen Vorerfahrungen der Versuchspersonen, beurteilt.

Vorhersagemodell für Gedächtnisaufgabe und ein *Query-by-humming*-Szenario:

Ein Ziel des hier beschriebenen Experimentes war es, aus empirischen Daten möglichst viele und möglichst weitreichende Erkenntnisse zum Wiedererkennensgedächtnis für Melodien zu gewinnen. Der zu diesem Zweck verwendete Forschungsansatz bestand darin, ein möglichst gutes bzw. genaues Vorhersagemodell für die Antworten der Versuchspersonen zu konstruieren. Die inhaltliche Interpretation des Zusammenspiels der verschiedenen Prädiktoren war ein zweiter Schritt, der erst nach der Festlegung des Vorhersagemodells getan wurde. Dies ist in den sogenannten *data-mining*-Ansätzen ein häufiges und erprobtes Vorgehen und unterscheidet sich deutlich vom klassischen Hypothesentesten: ‚Vorhersagequalität ist zunächst alles, eine Interpretation macht erst dann Sinn, wenn ein Modell mit guter Vorhersagequalität gefunden ist.‘

Ein dritter Schritt – der in dieser Arbeit jedoch nicht geleistet wird – könnte die qualitativ-quantitative Beschreibung eines Gedächtnismoduls sein, welches die Enkodierung, Speicherung und den Abruf von Melodien leistet. Ein Paradebeispiele für eine solche modulare Beschreibung, wie sie seit Fodor (1983) paradigmatisch für die kognitive Psychologie und auch die kognitive Musikforschung ist, ist beispielsweise das Modell des *Working Memory* von Alan Baddeley (1986). Für ein Modell dieser Art scheint jedoch die Datenlage im Bereich des Melodiegedächtnisses zu heterogen und insgesamt zu dünn.

Schließlich war auch das eingangs beschriebene *Qbh*-Szenario einmal mehr ein Leitbild. Dem Konstrukteur eines *Qbh*-System ist es in der Regel gleichgültig, welche Gedächtnis-komponenten bei der Benutzung seines *Qbh*-Systems zusammenspielen und wie diese Gedächtniskomponenten zu benennen sind. Wichtig ist für ihn, sein System so auf die Kapazitäten des menschlichen Gedächtnisses einzustellen, dass es einer Höchstzahl seiner Benutzer zufriedenstellende Ergebnisse liefert. Die konkreten Ergebnisse, die aus diesem Wiedererkennensexperiment für eine *Qbh*-Anwendung zu ziehen sind, sind u.a. folgende:

Den potentiellen Benutzern bleibt v.a. die durch *optil* gemessene Intervallfolge sowie die Akzentstruktur (*joint52*) der Melodien im Gedächtnis. Ein effektiver Algorithmus zur Bestimmung und zum Vergleich von melodischer Ähnlichkeit ist das in Kapitel 4 beschriebene Maß *optil* und für den Vergleich von Akzentstrukturen das in einem Exkurs dieses Kapitels konstruierte Maß *joint52*. Für das *matching* zwischen *query* und Referenzmelodie erscheint es deshalb sinnvoll, den akzentuierten Tönen einer Melodie ein besonderes Gewicht zu geben.

Die Wahrscheinlichkeit, dass dem potentiellen Benutzer Teile einer Melodie im Gedächtnis bleiben und damit, dass er sie in einer *Qbh*-Situation reproduziert, ist größer, je länger die Melodie (im Original) ist.

Potentielle Benutzer mit größerem Hintergrund an musikalischer Aktivität behalten Melodien besser im Gedächtnis. Wenn der Hintergrund des potentiellen Benutzers eingeschätzt werden kann, kann diese Information z.B. zum Einstellen eines Schwellwertes für Ähnlichkeitsvergleiche dienen.

Die konkrete rhythmische Gestaltung wird eher vergessen als die Abfolge von Intervallen oder Konturschritten, wie die signifikanten Unterschiede zwischen den Faktorstufen der Variable *Manipulationstyp* im vierfaktoriellen Modell mit den durch das Design kontrollierten Faktoren zeigten. Den Unterschieden im Rhythmus zwischen *query* und Referenzmelodie sollte also weniger Gewicht gegeben werden.

Besondere Bedeutung besitzt die Melodiekontur zudem bei der Erinnerung an Melodien, die früh im Verlauf eines Stückes zu hören sind. Die Melodiekontur wird gegenüber den einfachen Intervallgrößen auch dann besonders gut erinnert, wenn seit dem Hören der Referenzmelodie bereits mehr Zeit vergangen ist. Auch dies lässt sich anhand des vierfaktoriellen Modells belegen, genauer gesagt durch den entsprechenden signifikanten Interaktionseffekt zwischen *Position der Referenzmelodie* und *Manipulationstyp*.

Schließlich scheint es keine (eindeutigen) Hinweise darauf zu geben, dass beim Hören von Musik unter quasi ‚echten‘ Bedingungen Gedächtnisrepräsentationen in Abhängigkeit von der Position eines Melodieteils erstellt werden. Die Position der Töne in einer Phrase spielte dabei genauso wenig eine Rolle wie die Position einer einzelnen Phrase in der Gesamtmelodie. Für die Speicherung von Referenzmelodien in einem *Qbh*-System bedeutet dies, dass alle singbaren Melodien in ganzer Länge gespeichert werden müssen und es nicht sinnvoll ist, bestimmten Tönen beim *matching* von *query* zu Referenz ein besonderes Gewicht aufgrund ihrer Position zu geben.

Vergleich verschiedener Modellierungsansätze für Daten aus musikpsychologischen Experimenten:

Zwei deutlich verschiedene Ansätze der statistischen Modellierung wurden für die Analyse der Experimentaldaten hinsichtlich ihrer praktischen Handhabbarkeit und der Verwendbarkeit ihrer Ergebnisse miteinander verglichen. Die resultierende Datenlage ist in vieler Hinsicht typisch für musikpsychologische Experimente, wie oben ausgeführt wurde. Es zeigte sich, dass eine bestimmte Klasse von Modellen, die Klassifizierungs- und Regressionsbäume, mit diesen typischen Eigenheiten der Daten in musikpsychologischen Experimenten sehr gut zurecht kommen und relativ klare und gut interpretierbare Ergebnisse liefern.

Da Baummodelle in der Musikforschung insgesamt extrem selten als Erklärungsmodelle genutzt werden, soll dieses Kapitel auch zur Erschließung von neuen Möglichkeiten der statistischen Modellierung innerhalb der Musikforschung dienen.

Konstruktion von Vergleichsdaten und Prädiktoren für Reproduktionsexperimente:

Im Entwurf dieser Arbeit war zusätzlich zu dem in diesem Kapitel beschriebenen Wiedererkennensexperiment ein Experiment zur Reproduktion von Melodien aus dem Gedächtnis geplant. Aufgrund des überbordenden Umfangs dieser Arbeit bereits zum jetzigen Zeitpunkt, konnte dieses geplante Reproduktionsexperiment nicht mehr in dem hier gesteckten Rahmen realisiert werden. Es drängt sich jedoch geradezu auf, die in diesem Kapitel referierten Ergebnisse, die mit Hilfe der Wiedererkennensprozedur gefunden wurden, schließlich noch mit Erkenntnissen zur Reproduktion von Melodien aus dem Gedächtnis zu vergleichen. Zumindest ist also hier die Ausgangslage für die Formulierung von Hypothesen geschaffen worden. Noch wichtiger als die konkreten Vergleichsdaten könnten sich bei der Auswertung von Daten eines Reproduktionsexperimentes die hier geleisteten Arbeiten bei

der Konstruktion von Prädiktoren erweisen. Beispielweise dürfte das Maß für die Akzentähnlichkeit auch für die Analyse von menschlichen Wiedergaben von Melodien von größter Bedeutung sein. Ähnlich verhält es sich mit den Indexmaßen zum Musikkonsum und zur musikalischen Aktivität der Versuchspersonen, die in diesem Kapitel konstruiert wurden.

6 Fazit

Das Fazit dieser Arbeit kann vergleichsweise kurz gehalten werden. Denn die Ergebnisse der Literaturlaufarbeitung und der empirischen Teile sind jeweils zum Ende der einzelnen Kapitel in eigenen Zusammenfassungen dargestellt worden. Einen immer noch schnellen aber weit detaillierteren Überblick über die Ergebnisse dieser Arbeit als in diesem Fazit erhält man demnach durch die Lektüre der Abschnitte 3.1.4 (absolutes Gehör, Tongedächtnis, Gedächtnis für absolute Tonhöhen ganzer Musikstücke), 3.2.7 (Gedächtnis für kurze Tonfolgen), 3.3.5 (Gedächtnis für Melodien), 4.6 (eigene Experimente zur melodischen Ähnlichkeit) und 5.7 (eigenes Experiment zum Wiedererkennen von unbekannten Melodien).

Die Zusammenfassungen des Literaturkapitels (Kap. 3) zu den empirischen Forschungen zum Melodiegedächtnis wurden in Tabellenform dargestellt. So ist es möglich, sich relativ schnell einen Überblick über das empirische Wissen zum Gedächtnis für Einzeltöne und absolute Tonhöhen, zum Gedächtnis für kurze Tonfolgen und zum Gedächtnis für Melodien zu verschaffen. Es sollte jedoch bei der Benutzung der auf knappste Weise formulierten Ergebnisse in den Tabellen bedacht werden, dass diese Ergebnisse nur unter gewissen Bedingungen im Versuchslabor und unter Verwendung eines speziellen experimentellen Paradigmas zustande gekommen sind. Ob das angeführte Ergebnis auch für andere Hörbedingungen generalisierbar ist oder sogar für das Melodiegedächtnis in der ‚echten Welt‘ gilt, lässt sich in der Regel nicht im vorhinein entscheiden. Die Bedingungen, unter denen die empirischen Ergebnisse erzielt wurden, sind im Text des dritten Kapitels detailliert beschrieben. Diese Auflistung von empirischen Erkenntnissen zur Leistungsfähigkeit des menschlichen Melodiegedächtnisses bzw. zu dessen Störbarkeit und Fehlersystematik lässt sich sehr gut in Ergänzung zu den mehr theoretisch orientierten oder qualitativ zusammenfassenden Beschreibungen des Melodiegedächtnisses lesen, wie sie beispielsweise Berz (1995) und Snyder (2000) bzw. Crowder (1993), Bartlett (1996) oder Vazan (2001) vorgelegt haben. Direkte Gegenüberstellungen oder Auflistungen der empirischen Ergebnisse aus der Literatur fehlen in all diesen Überblickstexten.

Neben der durch die Tabellenform leicht fassbaren Aufbereitung der empirischen Ergebnisse ist in Kapitel 3 vor allem die methodische Diskussion der experimentellen Techniken bedeutsam, die in der Forschung zum Melodiegedächtnis besonders prominent vertreten

sind. Die Auseinandersetzung mit dem experimentellen Wiedererkennensparadigma und den für die Auswertung der experimentellen Daten häufig verwendeten Maßen d' und AUC aus der *Signal Detection Theory* zeigt die Grenzen auf, innerhalb derer – unter Verwendung dieser empirischen Ansätze – Erkenntnisse zum Melodiegedächtnis gewonnen werden können.

Das vierte Kapitel beschäftigt sich mit dem Thema der melodischen Ähnlichkeit. Hauptmotivation für die intensive Untersuchung verschiedener Ansätze zur Ähnlichkeitsmessung von Melodien ist die Konstruktion eines Messinstruments, welches für die Analyse von Leistungen des Melodiegedächtnisses eingesetzt werden kann. Für die Konstruktion dieses Messinstruments wird eine Vielzahl der in der einschlägigen Literatur diskutierten Ansätze zur Ähnlichkeitsmessung aufbereitet und für einen Abgleich mit im Experiment erhobenen Daten vergleichbar gemacht.

In drei Experimenten werden jeweils unter variierten Bedingungen Ähnlichkeitsurteile von Personen mit guter musikalischer Vorbildung erhoben. Es werden jedoch nur die Daten der Versuchspersonen zur weiteren Verarbeitung ausgewählt, die auf individueller Basis ein stabiles und konsistentes Ähnlichkeitsurteil zeigen. Im Ergebnis zeigt sich, dass die individuell stabilen und konsistenten Urteile auch intersubjektiv eine sehr hohe Übereinstimmung aufweisen. Das Mittel der menschlichen Ähnlichkeitsurteile kann somit als verlässliche ‚Eichgröße‘ für die Modellierung mit algorithmischen Maßen dienen.

Für die drei unterschiedlichen experimentellen Kontexte werden mittels linearer Regressionsanalysen drei verschiedene optimierte Ähnlichkeitsmodelle ermittelt, die aus einer Kombination von zwei bzw. drei verschiedenen Ähnlichkeitsalgorithmen bestehen. Für die hier erhobenen Daten erweisen sich diese Hybridmaße als ebenso robust wie effektiv. Keines der 24 getesteten algorithmischen Einzelmaße erreicht die selben guten Klassifizierungs- und Vorhersageleistungen wie die Hybridmaße.

Diese Hybridmaße können in gegebenen Anwendungssituationen, deren Randbedingungen bekannt sind, ohne weiteres eingesetzt werden. Zu diesen Anwendungssituationen gehören beispielsweise die Verwendung eines Ähnlichkeitsmaßes für das *matching* von *query* und Referenzmelodien in einem *Query-by-humming*-System, die Klassifizierung und Ordnung von großen Melodiedatenbank nach dem Ähnlichkeitskriterium (vgl. Müllensiefen & Frieler, 2004) oder die Analyse von aus dem Gedächtnis gesungenen Melodien in einem musikpsychologischen Experiment.

Neben der Konstruktion dieser Werkzeuge zur Ähnlichkeitsmessung wird in Kapitel 4 ein Modell des Ähnlichkeitsurteils der ausgewählten Versuchspersonen anhand der experimentellen Variablen aufgestellt. Auch dieses Modell ist das Ergebnis einer Serie von linearen Regressionsanalysen. Es weist sehr gute Modellwerte auf (Erklärung von 82% der Varianz in den experimentellen Daten; Korrelation der vorhergesagten mit den beobachteten Werten von $r = 0,907$) und erklärt das Urteil der Versuchspersonen als gewichtete Summe der Fehlerzahlen in den einzelnen manipulierten Dimensionen (Rhythmus, Intervalle, Kontur, Phrasenreihenfolge und ‚Modulation‘) sowie der Position und Dichte der Intervalle in der Melodie. Dies ist nach Wissen des Autors das erste Modell zur Ähnlichkeitswahrnehmung von Melodien, das den Großteil von experimentell erhobenen Ähnlichkeitsurteilen erklären kann.

Schließlich wird in Kapitel 4 noch der Raum der getesteten algorithmischen Ähnlichkeitsmaße durch ein multidimensionales Skalierungsmodell (MDS-Modell) dargestellt. Im Gegensatz zu den in der Literatur genannten Messdimensionen, zeigt sich, dass sich die unterschiedlichen Messresultate der einzelnen Maße nur durch zwei Dimensionen gut erklären lassen. Nach der Interpretation des MDS-Modells sind diese zwei Dimensionen erstens die Beachtung von globalen vs. lokalen Übereinstimmungen zwischen zwei Melodien und zweitens die Berücksichtigung vs. Nicht-Berücksichtigung von rhythmischer Information.

Das fünfte Kapitel beschäftigt sich mit der Konstruktion eines statistischen Modells zur Vorhersage von menschlichen Urteilen in einem Wiedererkennensexperiment. Aufgabe der musikalisch durchschnittlich vorgebildeten Versuchspersonen war es anzugeben, ob eine Vergleichsmelodie identisch oder leicht verändert in einem zuvor gehörten instrumentalen Musikstück vorhanden gewesen ist. Zur Vorhersage der Antworten der Versuchspersonen (identisch/verschieden-Entscheidung sowie Urteilssicherheit) werden ausschließlich vom experimentellen Design unabhängig messbare Größen verwendet. Zu diesen insgesamt 19 Größen zählen ein optimiertes Ähnlichkeitsmaß (aus Kapitel 4), ein Maß zur Bestimmung der Ähnlichkeit der Akzentstrukturen von Melodien (die Konstruktion dieses Maßes ist detailliert in einem Exkurs in 5.6.1.1 beschrieben), die Position der Vergleichsmelodie im Musikstück, die Länge der Vergleichsmelodie und der aktive wie passive musikalische Hintergrund der Versuchspersonen.

Zur statistischen Modellierung werden zwei grundsätzlich verschiedene Ansätze verwendet: Die lineare Modellbildung (Regression) sowie die Klassifizierungs- und Regressions-

bäume. Eine einfach nachvollziehbare methodische Beschreibung der in der Musikpsychologie extrem selten eingesetzten Baummodelle findet sich in Abschnitt 5.6.2.1.

Es zeigt sich, dass alle Modelle in etwa die gleiche Modellgüte erreichen, mit einem leichten Vorteil für den Regressionsbaum. Grob gesagt werden durch die Modelle etwa zwei Drittel der identisch/verschieden-Urteile richtig vorhergesagt und etwa ein Drittel der Werte auf der kombinierten, sechsstufigen Skala von identisch/verschieden-Urteil und Urteilssicherheit.

Inhaltlich zeigen die beiden Baummodelle große Übereinstimmungen und auch eine gewisse Ähnlichkeit mit dem linearen Modell. Für die Urteile der Versuchspersonen ist zunächst bedeutsam, ob es sich um (messbar) identische oder nicht-identische Melodien handelt. Bei nicht-identischen Melodien ist der Grad der Ähnlichkeit der Akzentstrukturen bzw. die melodische Ähnlichkeit (optimiertes Maß aus Kapitel 4) wichtig. Die Versuchspersonen sind also durchaus in der Lage, objektive Unterschiede und deren Ausprägung zwischen einer Melodie in einem Musikstück und einer Vergleichsmelodie festzustellen. Sind die im Gedächtnis zu vergleichenden Melodien identisch, machen die Versuchspersonen ihr Urteil von der Länge der Vergleichsmelodie abhängig. Längere Melodien werden generell mit einer höheren Urteilssicherheit als identisch bezeichnet. Dies mag dadurch erklärt werden, dass die Versuchspersonen mehr melodisches Material (Töne) in der Wiedererkennensphase angeboten bekommen, das sie mit eigenen Gedächtnisinhalten vergleichen können. Weiter ließe sich interpretieren, dass die Gedächtnisrepräsentationen von Melodien, die einmal innerhalb eines Musikstücks gehört werden, unvollständig sind. Je umfangreicher die Vergleichsmelodien sind, desto höher ist die Wahrscheinlichkeit, dass Teile von ihr wiedererkannt werden. Die Versuchspersonen scheinen also die wiedererkannten Melodieteile aufzusummieren und ab einem gewissen Schwellenwert für wiedererkannte Melodieteile das identisch-Urteil abzugeben. Wird dieser Schwellenwert nicht erreicht, z.B. weil die Vergleichsmelodie nur wenige Melodieteile enthält, so urteilen die Versuchspersonen eher, dass die Vergleichsmelodie nicht identisch zur Referenzmelodie gewesen ist.

Schließlich spielt noch die musikalische Aktivität der Versuchspersonen in der Vergangenheit eine Rolle für das Melodiegedächtnis. Personen mit geringerem musikalischen Training neigen eher dazu, kurze und identische Vergleichsmelodien als verschieden von der Referenzmelodie zu bezeichnen. Dies lässt den Schluss zu, dass das Melodiegedächtnis bei musikalisch aktiveren Personen leistungsfähiger ist, da bei ihnen auch das melodische

Material von kürzeren Melodien ausreicht, um den genannten Schwellenwert für das identisch-Urteil zu erreichen.

Weder das Aufsummieren noch der Schwellenwert für die Urteilsabgabe sind den Versuchspersonen bewusst. Diese Annahmen entstammen der Interpretation des Modells durch den Autor. Im Gegenteil verwenden die Versuchspersonen ganz unterschiedliche Gedächtnisstrategien zur Bewältigung der experimentellen Aufgabe, wie eine qualitative Befragung ergibt. Trotzdem mögen die quantitativ handhabbaren Modellannahmen des Aufsummierens und des Urteilsschwellenwertes hilfreiche Heuristiken z.B. für die Parametrisierung eines *Query-by-humming*-Systems oder der Analyse von Gedächtnisleistungen aus anderen wissenschaftlichen Experimenten sein.

Zwar präsentiert diese Arbeit eine ganze Reihe von neuen Erkenntnissen zum Melodiegedächtnis und zur menschlichen Wahrnehmung von melodischer Ähnlichkeit, die die existente Literatur in diesem Forschungsfeld bereichern dürften. Dennoch lässt sich auch eine zukünftige Fortführung der Ergebnisse dieser Arbeit in den Blick nehmen:

Die Vorhersagequalität der aufgestellten statistischen Erklärungsmodelle ist zwar etwa doppelt so gut wie die eines Zufallsmodells, aber immer noch nicht vollkommen zufriedenstellend. Wie eine grafische Inspektion und zugehörige nicht-parametrische Tests hinsichtlich der Verteilung der nicht durch den Regressionsbaum erklärten Anteile an den Urteilen (Residuen) zeigen, scheinen die 36 verwendeten Musikstücke und Testmelodien Eigenschaften zu besitzen, die nicht durch die verwendeten Prädiktoren gemessen werden und dennoch das Urteil der Versuchspersonen systematisch beeinflussen. In der Sprache des Statistikers: Die Residuen verteilen sich nicht normal.

Als bisher nicht beachtete Größen können die in vielen Studien zur musikalischen Merkmalsextraktion verwendeten Größen, wie Gängigkeit, Plausibilität oder Schema-Konformität, vermutet werden, die auf der Verteilung musikalischer oder melodischer Eigenschaften in größeren Musik- oder Melodiesammlungen beruhen. Dies scheint auch intuitiv sehr plausibel, denn die hier verwendeten Maße zur Übereinstimmung von musikalischen Merkmalen operieren ausschließlich auf dem Vergleich von je zwei Melodien und in Unkenntnis darüber, welche Unterschiede zwischen Melodien gewöhnlich oder banal und welche signifikant oder außergewöhnlich sind. Dies kann z.B. durch die Modellierung musikalischen Vorwissens anhand der statistischen Regelmäßigkeiten von Melodien eines bestimmten Stils geschehen.

Eine Fortführung dieser Arbeit müsste genau an diesem Punkt ansetzen. Das Vorkommen von bestimmten musikalischen Eigenschaften in Melodien eines bestimmten Repertoires, das den Versuchspersonen bekannt ist (z.B. die hier getestete Popmusik) muss anhand einer für dieses Repertoire repräsentativen Melodiedatenbank quantifiziert werden. Diese quantifizierten Merkmale können dann entweder als zusätzliche Prädiktoren in einem der getesteten Vorhersagemodelle verwendet werden oder die bereits verwendeten Prädiktoren können in Kenntnis der Häufigkeit der gemessenen Melodieeigenschaften neu geschätzt werden.

In jedem Fall sollte die Ausnutzung von Informationen zur Häufigkeitsverteilung von Melodiemerkmale zu einer wesentlich besseren Vorhersagequalität hinsichtlich der Gedächtnisleistungen in dem in dieser Arbeit durchgeführten Wiedererkennensexperiment führen. Aber auch für die Re-Analyse weiterer experimenteller Daten zum Melodiegedächtnis (z.B. Sloboda & Parker, 1985; Oura & Hatano, 1988) und zur Optimierung von *Music-Information-Retrieval*-Systemen, die die Variabilität und Konstanz von Melodien in der Erinnerung berücksichtigen müssen, scheint die in diesem Abschnitt skizzierte Fortführung dieser Arbeit sehr vielversprechend.

7 Literatur:

- Abraham, O. (1901). Das absolute Tonbewußtsein. *Sammelbände der Internationalen Musikgesellschaft*, 3, S. 1-86.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on automatic control*, Vol. AC-19, S. 716-723.
- Altenmüller, E. (1995). Gehör [Abschnitt: Absolutes Gehör]. In: *Musik in Geschichte und Gegenwart*, Sachteil. Kassel: Bärenreiter, Sp. 1101-1102.
- Anderson, J.A. (1995). *Learning and memory: an integrated approach*. New York: Wiley.
- Atkinson, R. & Shiffrin, R. (1968). Human memory: a proposed system and its control processes. In: Spence, K. & Spence, J. (Hg.), *The psychology of learning and motivation* (Vol. 2). New York: Academic Press.
- Attneave, F. & Olson, R.K. (1971). Pitch as a medium: a new approach to psychophysical scaling. *The American Journal of Psychology*, 84, S.147-166.
- Bachem, A. (1955). Absolute pitch. *Journal of the Acoustical Society of America*, 27, S. 1180-1185.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2003). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Zehnte, neu bearbeitete und erweiterte Auflage*. Berlin: Springer.
- Bäckman, L. & Herlitz, A. (1990). The relationship between prior knowledge and face recognition in normal aging and Alzheimer's disease. *Journal of Gerontology*, 45, S. 94-100.
- Baddeley, A. (1986). *Working memory*. Oxford: University Press.
- Baddeley, A. (1997). *Human memory. theory and practice*. Hove: Psychology Press.
- Baddeley, A. & Logie, R. (1992). Auditory Imagery and Working Memory. In: Reisberg, D. (Hg.), *Auditory Imagery*. Hillsdale (NJ): Lawrence Erlbaum, S. 179-197.
- Baharloo, S., Johnston, P.A., Service, S.K., Gitschier, J. & Freimer, N.B. (1998). Absolute pitch: An approach for identification of genetic and nongenetic components. *American Journal of Human Genetics*, 62, S. 224-231.
- Banks, W.P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, 74 (2), S. 81-99.
- Barlow, H & Morgenstern, S. (1949). *A dictionary of musical themes*. London: Ernest Benn.
- Baroni, M. & Callegari, L. (1984). Antiche canzoni francesi. Uno studio di metrica generativa. *Quaderni di informatica musicale*, 5.
- Baroni, M., Dalmonte, R. & Jacoboni, C. (1992). Theory and Analysis of European Melody. In: Marsden, A & Pople, A. (Hg.), *Computer Representations and Models in Music*. San Diego: Academic Press, S. 187-206.
- Bartlett, D.L. (1996). Tonal And Musical Memory. In: Hodges, D.A. (Hg.). *Handbook of Music Psychology*, 2nd edition. San Antonio: IMR Press, S. 177-195.
- Bartlett, F.C. (1932). *Remembering*. Cambridge: University Press.

- Bartlett, J.C., Halpern, A.R. & Dowling, W.J. (1995). Recognition of familiar and unfamiliar melodies in normal aging and Alzheimer's disease. *Memory & Cognition*, 23 (5), S. 531-546.
- Bartlett, J.C. & Snelus, P. (1980). Lifespan memory for popular songs. *American Journal of Psychology*, 93 (3), S. 551-560.
- Bartók, B. (1976 [1936]). Why and How Do We Collect Folk Music? In: Suchoff, B. (Hg), *Béla Bartók. Essays*. London: Faber & Faber, S. 9-24.
- Bartók, B. & Lord, A.B. (1951) *Serbo-Croatian Folk Songs: Texts and Transcriptions of Seventy-Five Folk Songs from the Milman Parry Collection and a Morphology of Serbo-Croatian Folk Melodies*. New York: Columbia University Press.
- Bergson, T.R. & Trehub, S.E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science*, 13, S. 72-75.
- Berlyne, D.E. (1971). *Aesthetics and psychobiology*. New York: Appelton-Century-Croft.
- Berlyne, D.E. (Hg.) (1974). *Studies in the New Experimental Aesthetics. Steps toward an objective Psychology of Aesthetic Appreciation*. Washington: Hemisphere Publishing.
- Berz, W.L. (1995). Working Memory in Music. A Theoretical Model. *Music Perception*, 12 (3), S. 353-364.
- Bharucha, J.J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, 5, S. 1-30.
- Bharucha, J.J. (1989). Pitch, harmony, and neural nets: A psychological perspective. *Computer Music Journal*, 13, S. 84-95.
- Bharucha, J.J. & Krumhansl, C. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, 13, 63-103.
- Bharucha, J.J. & Stoeckig, K. (1987). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance*, 12, S. 403-410.
- Bigand, E. (1990). Abstractions of two forms of underlying structure in a tonal melody. *Psychology of Music*, 18, S. 45-99.
- Bigand, E. (1993). Contributions of music to research on human auditory cognition. In: McAdams, S. & Bigand, E. (Hg.). *Thinking in Sound. The cognitive Psychology of Audition*. New York: Oxford University Press, S. 231-277.
- Bischoff, L.A., Granot, R.I. & Donchin, E. (2000). Recent discoveries in the psychophysiology of absolute pitch. In: Woods, C, Luck, G., Brochard, R., Seddon, F. & Sloboda, J.A. (Hg.), *Proceedings of the 6th International Conference on Music Perception and Cognition*, CD-ROM. Keele: University of Keele.
- Bod, R. (2002). Memory-Based Models of Melodic Analysis. Challenging the Gestalt Principles. *Journal of New Music Research*, 31 (1), S. 27-36.
- Boltz, M. (1991). Some structural determinants of melody recall. *Memory & Cognition*, 19 (3), S. 239-251.
- Boltz, M. (1992). The remembering of auditory event durations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, S. 938-956.

- Boltz, M. (1993). The generation of temporal and melodic expectancies during musical listening. *Perception & Psychophysics*, 53 (6), S. 585-600.
- Boltz, M. (1994). Changes in internal tempo and effects on the learning and remembering of event durations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, S. 1154-1171.
- Boltz, M. (1995). Effects of event structure on retrospective duration judgements. *Perception & Psychophysics*, 57, S. 1080-1096.
- Boltz, M. (1998). The Processing of Temporal and Nontemporal Information in the remembering of event Durations and Musical Structure. *Journal of Experimental Psychology: Human Perception and Performance*, 24 (4), S. 1087-1104.
- Boltz, M. (1999). The processing of melodic and temporal information: independent or unified dimensions? *Journal of New Music Research*, 28 (1), S. 67-79.
- Boltz, M. & Jones, M.R. (1986). Does rule Recursion Make Melodies Easier to reproduce? If Not, What Does?. *Cognitive Psychology*, 18, S. 389-431.
- Borg, I. & Staufenbiel, T. (1993). *Theorien und Methoden der Skalierung*. 2. revidierte Auflage. Bern: Huber.
- Bortz, J. (1999). *Statistik. Für Sozialwissenschaftler*. 5. Aufl. Berlin: Springer.
- Brady, P.T. (1970). Fixed-scale mechanism of absolute pitch. *Journal of the Acoustical Society of America*, 48, S. 883-887.
- Bransford, J.D. & Johnson, M.K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behaviour*, 11, S. 717-726.
- Bredenkamp, J. (1996). Grundlagen experimenteller Methoden. In: Erdfelder, E., Mausfeld, R. Meiser, Th. & Rudinger, G. (Hg.), *Handbuch Quantitative Methoden*. Weinheim: Beltz Psychologie Verlags Union, S. 37-46.
- Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984). *Classification and Regression Trees*. Belmont (CA): Wadsworth.
- Brosius, F. (1998). *SPSS 8.0: Professionelle Statistik unter Windows*. Bonn: MITP-Verlag.
- Brown, J.A. (1958). Some tests of decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10, S. 12-21.
- Bruhn, H. (1988). *Harmonielehre als Grammatik der Musik*. München: Psychologische Verlagsunion.
- Bruhn, H. (1991). Hör auf zu singen – Du bist ja ein Brummer! In: Kraemer, R.D. (Hg.), *Musikpädagogik. Unterricht – Forschung – Ausbildung (Karl Graml zum 70. Geburtstag)*. Mainz: Schott, S. 49-63.
- Bruhn, H. Rösing, H. (1993). Amateurmusiker. In: Bruhn, H. Oerter, R. & Rösing, H. (Hg.), *Musikpsychologie. Ein Handbuch*. Reinbek: Rowohlt, S. 221-228.
- Burge, D. (1981). *Perfect pitch: Color hearing for expanded musical awareness*. Fairfield (IA): American Educational Music Publications.
- Burge, D.L. (1986). *The Perfect Pitch Ear-Training Course*. American Educational Music Publications, Inc.
- Butler, D. & Ward, W.D. (1988). Effacing the memory of musical pitch. *Music Perception*, 5, S. 251-260.

- Buttram, J. (1996). Learning Theory and related Developments. Overview and Applications in Music Education and Music Therapy. In: *Handbook of Music Psychology*, 2nd edition. San Antonio: IMR Press, S. 401-467.
- Cambouropoulos, E. (2003). Musical Pattern Extraction for Melodic Structure. In: Kopiez, R., Lehmann, A.C., Walther, I. & Wolf, C. (Hg.), *Proceedings of the 5th triennial ESCOM Conference, Hannover 2003*. CD-ROM.
- Carlsen, J.C. (1981). Some factors which influence melodic expectancy. *Psychomusicology*, 1, S. 12-29.
- Carlsen, J.C. (1982). Musical expectancy: some perspectives. *Council for Research in Music Education*. 71, S. 4-14.
- Carroll, J.B. (1975). Speed and accuracy of absolute pitch judgements. Some latter day resultzs. *Educational Testing Service research Bulletin*, 35 (Princeton, NJ: ETS).
- Carterette, E.C., Kohl, D.C. & Pitt, M.A. (1986). Similarities among Transformed Melodies: The Abstraction of Invariants. *Music Perception*, 3 (4), S. 393-410.
- Chaffin, R. (2002). Bob Snyder, Music and Memory: An Introduction. *Music Perception*, 20 (1), S.101-106.
- Chiappe, P. & Schmuckler, M.A. (1997). Phrasing influences the recognition of melodies. *Psychonomic Bulletin & Review*, 4 (2), S. 254-259.
- Chin, Ch.S. (1997). The development of absolute pitch. In: Gabrielsson, A. (Hg.), *Proceedings of the Third Triennial ESCOM Conference*. Uppsala: Uppsala University, S. 105-110.
- Clarke, E.F. (1985). Some aspects of rhythm and expression in performances of Erik Satie's "Gnosienne No. 5". *Music Perception*, 2 (3), S. 299-328.
- Clarke, E.F. (1987). Categorical rhythm perception. An ecological persepective. In: Gabrielsson, A. (Hg.), *Action and perception in rhythm and music*. Stockholm: Royal Swedish Academy of Music, S. 19-34.
- Cohen, A.J., Trehub, S.E. & Thorpe, L.A. (1989). Effects of uncertainty on melodic information processing. *Perception & Psychophysics*, 46 (1), S. 18-28.
- Cohen, G. (1998). *Memory in the Real World*, 2nd Edition. Hove: Psychology Press.
- Craik, F. & Lockhart, R. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, 11, S. 671-684.
- Cronin, Ch. (1998). Concepts of Melodic Similarity in Music-Copyright Infringement Suits. In: Hewlett, W.B. & Selfridge-Field, E (Hg.), *Melodic Similarity. Concepts, Procedures, and Applications. Computing in Musicology 11*. Cambridge (MA): MIT Press, S. 187-209.
- Crowder, R.G. (1989). Imagery for Musical Timbre. *Journal of Experimental Psychology*, 15 (3), S. 472-478.
- Crowder, R.G. (1993). Auditory Memory. In: McAdams, S. & Bigand, E. (Hg.), *Thinking in sound. The cognitive Psychology of Human Audition*. Oxford: Clarendon Press. S. 113-145.
- Crowder, R.G., Serafine, M.L. & Repp, B.H. (1990). Physical interaction and assiciation by contiguity in memory for the words and melodies of songs. *Memory & Cognition*, 18 (5), S. 469-476.

- Crozier, J.B. (1997). Absolute pitch: Practice makes perfect, the earlier, the better. *Psychology of Music*, 25, S. 110-119.
- Cuddy, L.L. (1971). Absolute judgement of musically related pure tones. *Canadian Journal of Psychology*, 25, S. 42-55.
- Cuddy, L.L. & Cohen, A.J. (1976). Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology*, 28, S. 255-270.
- Cuddy, L.L., Cohen, A.J. & Mewhort, D.J.K. (1981). Perception of Structure in Short Melodic Sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 7 (4), S. 869-883.
- Cuddy, L.L. & Lyons, H.I. (1981). Musical Pattern Recognition: A Comparison of Listening to and Studying Structures and Tonal Ambiguities. *Psychomusicology*, 1 (2), S.15-33.
- Cutietta, R.A. & Booth, G.D. (1996). The Influence of Metre, Mode, Interval Type und Contour in Repeated Melodic Free-Recall. *Psychology of Music*, 18, S. 45-59.
- Dannenberg, R.B., Birmingham, W.P., Tzanetakis, G., Meek, C., Hu, N. & Pardo, B. (2004). The MUSART Testbed for Query-By-Humming Evaluation. *Computer Music Journal*, 28 (2), S. 34-48.
- Darwin, C.J., Turvey, M.T. & Crowder, R.G. (1972). An auditory analogue of the Sperling partial report procedure: evidence for brief auditory storage. *Cognitive Psychology* 3, S. 255-267
- Davies, J.B. & Yelland, A. (1977). Effects of two Trainig Procedures on the Production of Melodic contour, in short-Term Memory for Tonal Sequences. *Psychology of Music*, 5 (2), S. 3-9.
- Delis, D., Fleer, J. & Kerr, N.H. (1978). Memory for Music. *Perception & Psychophysics*, 23 (3), S. 215-218.
- Deutsch, D. (1969). Music recognition. *Psychological Review*, 76, S. 300-307.
- Deutsch, D. (1970). Tones and numbers. Specifity of interference in short-term memory. *Science*, 168, S. 1604-1605.
- Deutsch, D. (1972). Octave generalization and tune recognition. *Perception & Psychophysics*, 1972, 11 (6), S. 411-412.
- Deutsch, D. (1973). Octave generalization of specific interference effects in memory for tonal pitch. *Journal of Experimental Psychology*, 93, S. 156-162.
- Deutsch, D. (1974). Generality of inference by tonal stimuli in recognition memory for pitch. *Quarterly Journal of Experimental Psychology*, 26, S. 229-234.
- Deutsch, D. (1975a). The organization of short-term memory for a single acoustic attribute. In: Deutsch, D. & Deutsch, J.A. (Hg.), *Short term memory*. New York: Academic Press.
- Deutsch, D. (1975b). Two-channel listening to musical sclaes. *Journal of the Acoustical Society of America*, 57, 1156-1160.
- Deutsch, D. (1978). Delayed pitch comparisions and the principle of proximity. *Perception & Psychophysics*, 23, S. 227-230.
- Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, 28 (5), S. 381-389.

- Deutsch, D. & Boulanger, R.C. (1984). Octave equivalence and the processing of tonal sequences. *Music Perception*, 3, S. 40-51.
- Deutsch, D. & Feroe, J. (1975). Disinhibition in pitch memory. *Perception & Psychophysics*, 17, S. 320-324.
- Deutsch, D. & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review*, 88, S. 503-522.
- Deutsch, D., Henthorn, T. & Dolson, M. (2004). Absolute Pitch, Speech, and Tone Language: Some Experiments and a Proposed Framework. *Music Perception*, 21 (3), S. 339-356.
- Dewar, K.M., Cuddy, L.L. & Mewhort, D.J.K. (1977). Recognition Memory for Single Tones With and Without Context. *Journal of Experimental Psychology: Human Learning and Memory*, 3 (1), S. 60-67.
- DeWitt, L.A. & Crowder, R.G. (1986). Recognition of Novel Melodies after Brief Delays. *Music Perception*, 3 (3), S. 259-274.
- DeWitt, L.A. & Samuel, A.G. (1990). The role of Knowledge-Based Expectations in Music Perception: Evidence From Musical Restoration. *Journal of Experimental Psychology: General*, 119 (2), S. 123-144.
- Dowling, W.J. (1972). Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion. *Perception & Psychophysics*, 12 (5), S. 417-421.
- Dowling, W.J. (1973a). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics*, 14 (1), S. 37-40.
- Dowling, W.J. (1973b). The perception of interleaved melodies. *Cognitive Psychology*, 5, S. 322-337.
- Dowling, W.J. (1978). Scale and contour: Two Components of a Theory of Memory for Melodies. *Psychological Review*, 85 (4), S. 341-354.
- Dowling, W.J. (1984). Musical experience and tonal scales in the recognition of octave-scrambled melodies. *Psychomusicology*, 4, S. 13-32.
- Dowling, W.J. (1986). Context Effects on Melody Recognition: Scale-Step versus Interval Representations. *Music Perception*, 3(3), S. 281-296.
- Dowling, W.J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, 50 (4), S. 305-313.
- Dowling, W.J. & Bartlett, J.C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology*, 1, S. 30-49.
- Dowling, W.J. & Fujitani, D.S. (1971). Contour, Interval, and Pitch Recognition in Memory for Melodies. *The Journal of the Acoustical Society of America*, 49 (2, Part 2), S. 524-531.
- Dowling, W.J. & Harwood, D.L. (1986). Melody: Attention and Memory. In: Dowling, W.J. & Harwood, D.L. (Hg.), *Music Cognition*. Orlando: Academic Press, S. 124-152.
- Dowling, W.J. & Hollombe, A.W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception & Psychophysics*, 21 (1), S. 60-64.

- Dowling, W.J., Kwak, S. & Andrews, M.W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics*, 57 (2), S. 136-149.
- Dowling, W.J., Tillmann, B. & Ayers, D.F. (2002). Memory and the Experience of Hearing Music. *Music Perception*, 19 (2), S. 249-276.
- Downie, J.S. (1999). *Evaluating a simple approach to music information retrieval. Conceiving melodic n-grams as text*. Unpublished PhD Thesis, University of Western Ontario, London (Ontario), Kanada. Internet: http://music-ir.org/~jdownie/downie_thesis.pdf, April 2003.
- Drake, C. & Bertrand, D. (2001). The quest for universals in temporal music processing. In: Zatorre, R.J. & Peretz, I. (Hg.), *The Biological Foundations of Music*. New York: Academy of Sciences, S. 17-27.
- Drake, C., Dowling, W.J. & Palmer, C. (1991). Accent Structures in the Reproduction of Simple Tunes by Children and Adult Pianists. *Music Perception*, 8 (3), S. 315-334.
- Ebbinghaus, H. (1885). *Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie*. Amsterdam: Bonset (Nachdruck der Ausgabe Leipzig 1885).
- Edelmann, W. (1996). *Lernpsychologie*. 5. Aufl. Weinheim: Beltz, Psychologische Verlags Union.
- Edworthy, J. (1983). Towards a contour-pitch continuum theory of memory for melodies. In: Rogers, D. & Sloboda, J.A. (Hg.), *The Acquisition of Symbolic Skills*. New York: Plenum, S. 263-271.
- Edworthy, J. (1985). Interval and Contour in Melody Processing. *Music Perception*, 2 (3), S. 375-388.
- van Egmond, R., Povel, D.J. & Maris, E. (1996). The influence of height and key on the perceptual similarity of transposed melodies. *Perception & Psychophysics*, 58, S. 1252-1259.
- Eerola, T., Järvinen, T., Louhivuori, J. & Toiviainen, P. (2001). Statistical Features and Perceived Similarity of Folk Melodies. *Music Perception*, 18 (3), S. 275-296.
- Eich, E. & Metcalfe, J. (1989). Mood dependent memory for internal versus external events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15 (3), S. 443-455.
- Eiting, M.H. (1984). Perceptual Similarities between Musical Motifs. *Music Perception*, 2 (1), S. 78-94.
- Feierabend, J.M., Saunders, T.C., Holahan, J.M. & Getnick, P. (1998). Song recognition among Preschool-Age children: An Investigation of Words and Music. *Journal of Research in Music Education*, 46 (3), S. 351-359.
- Fodor, J.A. (1983). *The modularity of mind*. Cambridge (MA): MIT Press.
- Frieler, K. (2004). Beat and meter extraction with gaussified onsets. *ISMIR 2004 Conference Proceedings. Fifth International Conference on Music Information Retrieval*. Barcelona, in press.
- Frieler, K. (2005). *Mathematische Musikanalyse. Theorie und Praxis*. Dissertation Universität Hamburg, in Arbeit.
- Gardiner, J.M., Kaminska, Z., Dixon, M. & Java, R.I. (1996). Repetition of previously novel melodies sometimes increases both remember and know responses in recognition memory. *Psychonomic Bulletin & Review*, 3 (3), S. 366-371.

- Gathercole, S. & Baddeley, A. (1993). *Working memory and language*. Hillsdale (NJ): Lawrence Erlbaum Associates.
- Gordon, E. (1986). A factor analysis of the Musical Aptitude Profile, the Primary Measures of Music Audiation, and the Intermediate Measures of Music Audiation. *Bulletin of the Council for Research in Music Education*, 87, S. 17-25.
- Green, D.M. & Moses, F.L. (1966). On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin*, 66 (3), S. 228-234.
- Green, D.M. & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Gregersen, P.K., Kowalsky, E., Kohn, N. & Marvin, E.W. (1999). Absolute pitch: Prevalence, ethnic variation, and estimation of the genetic component. *American Journal of Human Genetics*, 65, S. 911-913.
- Gribskov, M. & Devereux, J. (1991). *Sequence Analysis Primer*. New York: Stockton Press.
- Grote, K. (2001). Priming. In: Pethes, N & Ruchatz, J. (Hg.), *Gedächtnis und Erinnerung. Ein interdisziplinäres Lexikon*. Reinbek: Rowohlt, S. 457.
- Guido von Arezzo (o.J.). Prologus In Antiphonarum. In: *Tres Tractatuli Guidonis Aretini. Guidonis "Prologus In Antiphonarum"*. Divitiae Musicae Artis (DMA). A. III. Hg.: Joseph Smits van Waesberghe. Buren, Frits Knuf, 1975.
- Gusfield, D. (1997). *Algorithms on Strings, Trees, and Sequences. Computer Science and Computational Biology*. Cambridge: University Press.
- Hahn, K. (2002a). Absolute Memory for Melody: A study with choir singers. In: Stevens, C., Burnham, D., McPerson, G., Schubert, E. & Renwick, J. (Hg.), *Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney 2002*, CD-ROM. Adelaide: Casual Productions.
- Hahn, K. (2002b). Absolutes Gehör und das Erinnern von Melodien. In: Bullerjahn, C. (Hg.). *Stimme und Singen – Psychologische Aspekte. Tagungsband zur Jahrestagung der DGM an der Hochschule Magdeburg-Stendal*. Hannover: Eigenverlag., S. 25.
- Halpern, A. (1984). Organization in Memory for Familiar Songs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10 (3), S. 496-512.
- Halpern, A. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition*, 17 (5), S. 572-581.
- Halpern, A. Bartlett, J.C. & Dowling, W.J. (1995). Aging and expertise in the perception of musical transposition. *Psychology & Aging*, 10, S. 325-342.
- Handel, S. (1989). *Listening. An Introduction to the Perception of Auditory Events*. Cambridge (MA): MIT Press.
- Hantz, E. (1984). Studies in Musical Cognition: Comments from a Music Theorist. *Music Perception*, 2 (2), S. 245-264.
- Hart, J.T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, 56, S. 208-216.
- Hébert, S. & Peretz, I. (1997). Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? *Memory & Cognition*, 25 (4), S. 518-533.

- Hemker, Th. & Müllensiefen, D. (Hg.) (1997). *Medien-Musik-Mensch. Neue Medien und Musikwissenschaft*. Hamburg: von Bockel.
- Heyde, E.-M., (1987). *Was ist absolutes Hören? Eine musikpsychologische Untersuchung*. München: Profil.
- Hintzman, D.L. (1994). On explaining the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, S. 275-289.
- Hirsh, I.J, Monahan, C.B., Grant, K.W. & Singh, P.G. (1990). Studies in auditory timing: 1. Simple patterns. *Perception & Psychophysics*, 47, 3, S. 215-226.
- Hofmann-Engl, L. (2001). Towards a cognitive model of melodic similarity. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*. Internet: <http://ismir2001.indiana.edu/üpdf/hofmann-engl.pdf>, im Juni 2003.
- Hofmann-Engl, L. (2002). Rhythmic Similarity: A theoretical and empirical approach. In: Stevens, C., Burnham, D., McPerson, G., Schubert, E. & Renwick, J. (Hg.), *Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney 2002*, CD-ROM. Adelaide: Casual Productions.
- Holtzman, S. (1977). A program for key determination. *Interface*, 6 (1), S. 29-56.
- Hosmer, D.W. & Lemeshow, S. (2000). *Applied Logistic Regression*. Wiley, New York.
- Hunt, E.B., Marin, J. & Stone, P.J. (1966). *Experiments in induction*. New York: Academic Press.
- Hurni-Schlegel, L. & Lang, A. (1978). Verteilung, Korrelate und Veränderbarkeit der Tonhöhen-Identifikation (sog. Absolutes Musikgehör). *Schweizerische Zeitschrift für Psychologie*, 37, S. 265-292.
- Huron, D. & Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology*, 12, S. 152-169.
- Hyman, I.E. & Rubin, D.C. (1990). Memorybeatlia: A naturalistic study of long-term memory. *Memory & Cognition*, 18 (2), S. 205-214.
- Idson, W.L. & Massaro, D.W. (1976). Cross-octave masking of single tones and musical sequences: The effects of structure on auditory recognition. *Perception & Psychophysics*, 19 (2), S. 155-175.
- Idson, W.L. & Massaro, D.W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception & Psychophysics*, 24 (6), S. 551-565.
- Irtel, H. (1996). Methoden der Psychophysik. In: Erdfelder, E., Mausfeld, R. Meiser, Th. & Rudinger, G. (Hg.), *Handbuch Quantitative Methoden*. Weinheim: Beltz Psychologie Verlags Union, S. 479-489.
- Jesser, B. (1990). *Interaktive Melodieanalyse: Methodik und Anwendung computergestützter Analyseverfahren in Musikethnologie und Volksliedforschung. Typologische Untersuchung der Balladensammlung des DVA*. Bern: Peter Lang.
- Jones, M.R. (1981). Music as a stimulus of psychological motion. Part I. Some determinants of expectancies. *Psychomusicology*, 1, 34-51.
- Jones, M.R. (1982). Music as a stimulus of psychological motion. Part II. An expectancy model. *Psychomusicology*, 2, 1-13.

- Jones, M.R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception & Psychophysics*, 41, S. 621-634.
- Jones, M.R. (1993). Dynamics of Musical Patterns: How Do Melody and Rhythm Fit Together?. In: Tighe, Th. & Dowling, W.J. (Hg.). *Psychology and music: The understanding of melody and rhythm*. Hillsdale (NJ): Lawrence Erlbaum, 1993. S. 67-92.
- Jones, M.R. & Boltz, M. (1989). Dynamic Attending and Responses to Time. *Psychological Review*, 96 (3), S.459-491.
- Jones, M.R., Moynihan, H., MacKenzie, N. & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, 14 (4), S. 313-319.
- Jones, M.R. & Ralston J.T. (1991). Some influences of accent structure on melody recognition. *Memory & Cognition*, 19, S. 8-20.
- Jones, M.R., Summerell, L. & Marshburn, E. (1987). Recognizing Melodies: A Dynamic Interpretation. *The Quarterly Journal of Experimental Psychology*, 39A, S. 89-121.
- Jörger, Th.M. (1992). *Das Plagiat in der Populärmusik*. Baden-Baden: Nomos Verlagsgesellschaft.
- Jugendwerk der Deutschen Shell (Hg.) (2000). *Jugend 2000 – 13. Shell Jugendstudie*. Opladen: Leske + Budrich.
- Juhász, Z. (2000). A Model of Variation in the Music of a Hungarian Ethnic Group. *Journal of New Music Research*, 29 (2), S. 159-172.
- Juszyk, P.W. & Krumhansl, C. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of experimental Psychology: Human Perception and Performance*, 19, S. 627-640.
- Kallmann, H.J., Cameron, P.A., Beckstead, J.W. & Joyce, E. (1987). Ear of input as a determinant of pitch-memory interference. *Memory & Cognition*, 15, S. 454-460.
- Kauffman, W.H. (1990). Developmental Memory for Intact Music. *Psychomusicology*, 9 (1), S. 107-109.
- Kauffman, W.H. & Carlsen, J.C. (1989). Memory for Intact Music Works: The Importance of Music Expertise and Retention Interval. *Psychomusicology*, 8 (1), S. 3-20.
- Keller, K. & Rabson, C. (1980). *National tune index. 18th century secular music*. New York: University Music Education.
- Klinke, R., Boerger, G. & Gruber, J. (1970). The influence of the frequency relation in dichotic stimulation upon the cochlear nucleus activity. In: Plomp, R. Smoorenburg, G.F. (Hg.), *Frequency analysis and periodicity detection in hearing*. Leiden: Sijthoff.
- Kluge, R. (1974). *Faktorenanalytischen Typenbestimmung an Volksliedmelodien*. Leipzig: VEB Deutscher Verlag für Musik.
- Kluge, R. (1996). Ähnlichkeitskriterien für Melodieanalyse. *Systematische Musikwissenschaft*, 4 (1-2), S. 91-99.
- Kluwe, R. (1990). Gedächtnis und Wissen. In: H. Spada (Hg.), *Allgemeine Psychologie*. Stuttgart: Hans Huber, S. 115-187.

- Kluwe, R. (1996). Gedächtnis. *Wörterbuch der Kognitionswissenschaft*. Strube, G., Freksa, Ch. Hahn, U., Opwis, K. & Palm, G. (Hg.). Stuttgart: Klett-Cotta. S. 195-209.
- Kochmann, R. (1923). Über musikalische Gedächtnisbilder. *Zeitschrift für angewandte Psychologie*, 22, S. 329-351.
- Kötter, E. (1991). Zur Problematik des Reizmaterials in Experimenten der Musikpsychologie. Behne, K.-E., Jost, E. Kötter, E. & de la Motte-Haber, H. (Hg.). *Musikwissenschaft als Kulturwissenschaft. Festschrift zum 65. Geburtstag von H.-P. Reinecke*. Regensburg: Bosse, S. 187-196.
- Kötter, E. (1992a). Was ist musikalische Substanz? Zur Problematik des Reizmaterials in Hörversuchen zur melodischen Abstraktion. *Musikpädagogische Forschung*, 13, Essen: Verlag Blaue Eule, S. 138-145.
- Kötter, E. (1992b). Zur Prototypenbildung bei der Abstrahierung melodischen Materials. *Jahrbuch der Deutschen Gesellschaft für Musikpsychologie*, 9, S. 55-64.
- Krumhansl, C. (1983). Perceptual structures for tonal music. *Music Perception*, 1 (1), S. 28-62.
- Krumhansl, C. (1990). *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.
- Krumhansl, C. (1991). Memory for musical surface. *Memory & Cognition*, 19 (4), S. 401-411.
- Krumhansl, C. & Jusczyk, P.W. (1990). Infants' perception of phrase structure in music. *Psychological Science*, 1, S. 70-73.
- Krumhansl, C. & Kessler, E.J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89, S. 334-368.
- Krumhansl, C. & Toiviainen, P. (2001). Tonal Cognition. In: Zatorre, R.J. & Peretz, I. (Hg.), *The Biological Foundations of Music*. New York: Academy of Sciences, S. 77-91.
- Kruskal, J.B. & Wish, (1978). *Multidimensional scaling*. Beverly Hills: Sage.
- Kühn, C. (1987). *Formenlehre der Musik*. 6. Aufl. München u. Kassel: dtv / Bärenreiter.
- Lamont, A. & Dibben, N. (2001). Motivic Structure and the Perception of Similarity. *Music Perception*, 18 (3), S. 245-274.
- Large, E.W. & Jones, M.R. (1999). The dynamics of attending: How people track time varying events. *Psychological Review*, 106, S.119-159.
- Large, E.W., Palmer, C. & Pollack, J.B. (1995). Reduced Memory Representations for Music. *Cognitive Science*, 19, S. 53-96.
- Lehrdahl, F. & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge (MA): MIT Press.
- Leman, M. (2000). An Auditory Model of the Role of Short-Term Memory in Probe-Tone Ratings. *Music Perception*, 17 (4), S. 481-509.
- Leppig, M. (1987a). Musikuntersuchungen in Rechenautomaten. *Musica*, 41 (2), S. 140-150.
- Leppig, M. (1987b). Tonfolgenverarbeitungen in Rechenautomaten. Muster und Formen. *Zeitschrift für Musikpädagogik*, 42, S. 59-65.

- Levitin, D.J. (1994). Absolute memory for musical pitch: evidence from the production of learned melodies. *Perception & Psychophysics*, 56 (4), S. 414-423.
- Levitin, D.J. (1997). Pitch memory and the A440 scale. In: Gabrielsson, A. (Hg.), *Proceedings of the Third Triennial ESCOM Conference*. Uppsala: Uppsala University, S. 117-121.
- Levitin, D.J. (1999). Memory for Musical Attributes. In: Cook, P.R. (Hg.), *Music, Cognition, and Computerized Sound. An Introduction to Psychoacoustics*. Cambridge (MA): MIT Press, 1999. S. 209-227.
- Lindsay, A.T. (1996). *Using Contour as a mid-level Representation of Melody*. Unveröffentlichte Master-of-Science-Arbeit am Massachusetts Institute of Technology.
- Logie, R. H., & Edworthy, J. (1986). Shared mechanisms in the processing of verbal and musical material. In: Russell, D.G., Marks, D. & Richardson, J. (Hg.), *Imagery II*. Dunedin (New Zealand): Human Performance Associates, S. 33-37.
- Long, P.A. (1977). Relationships between pitch memory in short melodies and selected factors. *Journal of Research in Music Education*, 25 (4), S. 272-282.
- Longuet-Higgins, H. C. and Steedman, M. J. (1971). On interpreting bach. In: Meltzer, B. & Michie, D. (Hg.), *Machine Intelligence, Vol. 6*. Edinburgh: University Press, S. 221-241.
- Lopresti, D. & Wilfong, G. (1999). Cross-Domain Approximate String Matching. *Paper presented at the 6th International Symposium on String Processing and Information retrieval, Cancún, September 1999*. Internet: <http://www.bell-labs.com/org/1133/Research/DocumentAnalysis/papers/xdomain.pdf>, April 2003.
- MacMillan, N.A. & Creelman, C.D. (1991). *Detection Theory. A user's guide*. Cambridge: University Press.
- Martin, R.C., Wogalter, M.S. & Forlano, J.G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language*, 27, S. 382-398.
- Massaro, D.W., Kallman, H.J. & Kelly, J.L. (1980). The Role of Tone Height, Melodic Contour, and Tone Chroma in Melody Recognition. *Journal of Experimental Psychology: Human Learning and Memory*, 6 (1), S. 91-105.
- McAdams, S. & Matzkin, D. (2001). Similarity, Invariance, and Musical Variation. In: Zatorre, R.J. & Peretz, I (Hg.), *The Biological Foundations of Music*. New York: Academy of Sciences, S. 62-74.
- McNab, R.J., Smith, L.A., Witten, I.H., Henderson, C.L. & Cunningham, S.J. (1996). Towards the Digital Music Library: Tune retrieval from Acoustic Input. In: *Digital Libraries '96. Proceedings ACM Digital Libraries Conference*, S. 11-18.
- Meek, C. & Birmingham, W. (2002). Johnny Can't Sing. A Comprehensive Error Model for In: Fingerhut, M. (Hg.), *ISMIR 2002 Conference Proceedings. Third International Conference on Music Information Retrieval*. Paris: IRCAM, S. 124-132.
- Meinz, E.J. & Salthouse, T.A. (1998). The effects of age and experience on memory for visually presented music. *Journal of Gerontology: Psychological Sciences*, 53B (1), S. 60-69.
- Meiser, Th. (1997). *Arbeitsgedächtnis und Changing-State-Hypothese*. Dissertation Universität Heidelberg.

- Meyer, L.B. (1956). *Emotion and meaning in music*. Chicago: University Press.
- Meyer, L.B. (1973). *Explaining music. Essays and explorations*. Chicago: University Press.
- Mikumo, M. (199). Encoding Strategies for Tonal and Atonal Melodies. *Music Perception*, 10 (1), S. 73-82.
- Mikumo, M. (1994). Motor Encoding Strategy for Pitches of Melodies. *Music Perception*, 12 (2), S. 175-197.
- Miller, G. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63, S. 81-97.
- Miyazaki, K. (1988). Musical pitch identification by absolute pitch possessors. *Perception & Psychophysics*, 44, S. 501-512.
- Moles, A. (1971). *Informationstheorie und ästhetische Wahrnehmung*. Köln.
- Monahan, C.B. (1993). Parallels Between Pitch and Time and How They Go Together. In: Tighe, T.J. & Dowling, W.J. (Hg.), *Psychology and music: The understanding of melody and rhythm*. Hillsdale (NJ): Lawrence Erlbaum, S. 121-154.
- Monahan, C.B. & Carterette, E.C. (1985). Pitch and duration as determinants of musical space. *Music Perception*, 3 (1), S. 1-32
- Monahan, C.B. & Hirsh, I.J. (1990). Studies in auditory timing. 2. Rhythm patterns. *Perception & Psychophysics*, 47 (3), S. 227-242.
- Monahan, C.B., Kendall, R.A. & Carterette, E.C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception & Psychophysics*, 41 (6), S. 576-600.
- Mongeau, M. & Sankoff, D. (1990). Comparison of Musical Sequences. *Computers and the Humanities*, 24, S. 161-175
- Morgan, J.N. & Sonquist, J.A. (1963). Problems in the analysis of survey data and a proposal. *Journal of the American Statistical Association*, 58, S. 415-434.
- Morrow, D.G., Leirer, V.O., & Altieri, P.A. (1992). Aging, expertise, and narrative processing. *Psychology & Aging*, 7, S. 376-388.
- Morrow, D.G., Leirer, V.O., Altieri, P.A. & Fitzsimmons, C. (1994). When expertise reduces age differences in performance. *Psychology & Aging*, 9, S. 134-148.
- de la Motte, D. (1976). *Harmonielehre*. Kassel: Bärenreiter.
- de la Motte, D. (1990). *Musikalische Analyse. Mit kritischen Anmerkungen von Carl Dahlhaus*. 6. Aufl. Kassel: Bärenreiter.
- Müllensiefen, D. (1999). Radikaler Konstruktivismus und Musikwissenschaft: Ideen und Perspektiven. *Musicae Scientiae*, 3 (1), S. 95-116.
- Müllensiefen, D. (2000a). *Gedächtnisleistungen bei Hintergrundmusik*. Unveröffentlichte Magisterschrift, Universität Hamburg.
- Müllensiefen, D. (2000b). Die Wirkungen von Hintergrundmusik auf das Gedächtnis. *Systematische Musikwissenschaft*, 7 (1-2), S. 33-58.
- Müllensiefen, D. & Frieler, K. (2004). Cognitive Adequacy in the Measurement of Melodic Similarity: Algorithmic vs. Human Judgements. *Computing in Musicology*, 13, im Druck.

- Münzer, S. (1998). *Artikulatorische und tonale Unterdrückung in einer verbalen ISR-Aufgabe bei MusikerInnen und NichtmusikerInnen*. Unveröffentlichte Diplomarbeit in der Fachrichtung Psychologie. Saarbrücken: Universität des Saarlandes.
- Narmour, E. (1989). The „genetic code“ of melody: Cognitive structures generated by the implication-realization model. *Contemporary Music Review*, 4, S. 45-63.
- Narmour, E. (1990). *The analysis and cognition of melodic complexity*. Chicago: University Press.
- Nauck-Börner, Ch. (1988). Strukturen des musikalischen Gedächtnisses. Anmerkungen zu formalen Modellen der Repräsentation. In: Behne, K.-E., Kleinen, G., de la Motte-Haber, H. (Hg.), *Jahrbuch der Deutschen Gesellschaft für Musikpsychologie*, 5. Göttingen: Hogrefe, S. 55-66.
- Niketta, R. (1990). Was ist prototypische Rockmusik? Zum Zusammenhang zwischen Prototypikalität, Komplexität und ästhetischem Urteil. *Musikpsychologie. Jahrbuch der deutschen Gesellschaft für Musikpsychologie*, 7, S. 35-60.
- Norman, D.A. & Wickelgren, W.A. (1965). A short-term recognition memory for single digits and pairs of digits. *Journal of Experimental Psychology*, 70, S. 479-489.
- Novatris (2003). *NetCibles: Deutschland Erhebungswelle April 2003*. Unveröffentlichte Internetbefragung.
- Ogawa, Y., Kimura, T. & Mito, H. (1995). Modification of Musical Schema for Japanese Melody: A Study of Comprehensible and Memorable Melody. *Bulletin of the Council for Research in Music Education*, 127, S. 136-141.
- Olson, R.K. & Hanson, V. (1977). Interference effects in tone memory. *Memory & Cognition*, 5, S. 32-40.
- Ortmann, O. (1933). Some tonal determinants of melodic memory. *The Journal of Educational Psychology*, 24, S. 454-467.
- Oura, Y. & Hatano, G. (1988). Memory for Melodies among Subjects Differing in Age and Experience in Music. *Psychology of Music*, 16, S. 91-109.
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, 76, S. 241-263.
- Palisca, C. (1980). Guido of Arezzo. In: Sadie, S. (Hg.), *The New Grove Dictionary of Music and Musicians*. London: MacMillan Publisher Ltd.
- Palmer, Caroline & Krumhansl, Carol L. 1990. Mental Representations for Musical Meter. *Journal of Experimental Psychology: Human Perception and Performance* 1990, Vol. 16, No. 4, S. 728-741.
- Papez, J.W. (1937). A proposed mechanism of emotion. In: *Archives of Neurology and Psychiatry*, 79, S. 217-224.
- Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, 11, S. 409-464.
- Parsons, D. (1975). *The directory of tunes and musical themes*. New York: Spencer Brown.
- Pembroke, R. (1986). Interference of the transcription process and other selected variables on perception and memory during melodic dictation. *Journal of Research in Music Education*, 34 (4), S. 238-261.

- Pembrook, R. (1987). The effect of vocalization on melodic memory conservation. *Journal of Research in Music Education*, 35 (3), S. 155-70.
- Pauws, S. (2002). Cuby hum: A Fully Operational Query by Humming System. *ISMIR 2002 Conference Proceedings*. Paris: IRCAM, S. 187-196.
- Pechmann, Th. & Mohr, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory & Cognition*, 20 (3), S. 314-320.
- Pethes, N & Ruchatz, J. (2001). *Gedächtnis und Erinnerung. Ein interdisziplinäres Lexikon*. Reinbek: Rowohlt.
- Peynircioğlu, Z.F. Tekcan, A.I., Wagner, J.L., Baxter, T.L. & Shaffer, S.D. (1998). Name or hum that tune: Feeling of knowing for music. *Memory & Cognition*, 26 (6), S. 1131-1137.
- Piston, W. (1978). *Harmony. 4th Edition*. New York: W.W. Norton and Co.
- Polhardt-Gott, L. (1983). Emergence of thematic concepts in repeated listening to music. *Cognitive Psychology*, 15, S. 66-94.
- Pollack, I. & Hsieh, R. (1969). Sampling variability of the area under the ROC-curve and of d' . *Psychological Bulletin*, 71 (3), S. 161-173.
- Povel, D.J. (1981). Internal representations of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 7 (1), S. 3-18.
- Povel D.J. & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, 2 (4), 411-440.
- Povel, D.J. & Okkerman, H. (1981). Accents in Equitone Sequences. *Perception & Psychophysics*, 30 (6), S. 565-572.
- Profita, J. & Bidder, T.G. (1988). Perfect pitch. *American Journal of Medical Genetics*, 29, S. 763-771.
- Quinlan, J.R. (1979). Discovering rules by induction from large collections of examples. In: Michie, D. (Hg.), *Expert systems in the micro electronic age*. Edinburgh: University Press.
- Radvansky, G.A., Fleming, K.J. & Simmons, J.A. (1995). Timbre Reliance in Nonmusicians' and Musicians' Memory for Melodies. *Music Perception*, 13 (2), S. 127-140.
- Radvansky, G.A. & Potter, J.K. (2000). Source cuing: Memory for melodies. *Memory & Cognition*, 28 (5), S. 693-699.
- Rainey, D.W. & Larsen, J. (2002). The Effect of Familiar Melodies on Initial Learning and Long-term Memory for Uncennected Text. *Music Perception*, 20 (2), S. 173-186.
- Rakowski, A. & Morawska-Büngeler, M. (1987). In search for criteria of absolute pitch. *Archives of Acoustics*, 12, S. 75-87.
- von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung*, 18, S. 299-342.
- Révész, G. (1946). *Einführung in die Musikpsychologie*. Bern: A. Francke.
- Roberts, L.A. (1986). Modality and Suffix Effects in Memory for Melodic and Harmonic Musical Materials. *Cognitive Psychology*, 18, S. 123-157.

- Roberts, L.A., Millen, D.R., Palmer, C. & Tartter V.C. (1983). Modality and suffix effects in memory for music. (abstract). *Journal of the Acoustical Society of America*, Supplement 1, 74, S. S22-S23.
- Rösing, H. (1998). Musikgebrauch im täglichen Leben. In: Bruhn, H. & Rösing, H. (Hg.), *Musikwissenschaft. Ein Grundkurs*. Reinbek: Rowohlt, S. 107-129.
- Rösing, H. Bruhn, H. (1993). Typologie der Musikhörer. In: Bruhn, H. Oerter, R. & Rösing, H. (Hg.), *Musikpsychologie. Ein Handbuch*. Reinbek: Rowohlt, S. 130-136.
- Rosner, B.S. & Meyer, L.B. (1982). Melodic processes and the perception of music. In: Deutsch, D. (Hg.), *The psychology of music*. New York: Academic Press, S. 317-341.
- Rosner, B.S. & Meyer, L.B. (1986). The perceptual roles of melodic process, contour, and form. *Music Perception*, 4, S. 1-40.
- Russell, P.A. (1987). Memory for music: A study of musical and listener factors. *British Journal of Psychology*, 78, S. 335-347.
- Sakakibara, A. (1999). A longitudinal study of a process for acquiring absolute pitch. *Japanese Journal of Educational Psychology*, 47, S. 19-27.
- Sakakibara, A. (2000). A longitudinal study of the acquisition process of absolute pitch: An effect of subjects's age on the process. In: Woods, C, Luck, G., Brochard, R., Seddon, F. & Sloboda, J.A. (Hg.), *Proceedings of the 6th International Conference on Music Perception and Cognition*, CD-ROM. Keele: University of Keele.
- Salamé, P. & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology* 41 (A), S. 107-122.
- Salthouse, T.A. (1984). Effects of age and skill on typing. *Journal of Experimental Psychology: General*, 13, S. 345-371.
- Schacter, D.L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, S. 501-518.
- Schenker, H. (1932). *Fünf Urfurien-Tafeln*. Wien.
- Schmidt, S.J. (1991). Gedächtnisforschungen. Positionen, Probleme, Perspektiven. In: Schmidt, S.J. (Hg.), *Gedächtnis. Probleme und Perspektiven der interdisziplinären Gedächtnisforschung*. Frankfurt a.M.: Suhrkamp, S. 9-55.
- Schmuckler, M.A. (1989). Expectation in Music: Investigation of Melodic and Harmonic Processes. *Music Perception*, 7 (2), S. 109-150.
- Schmuckler, M.A. (1990). The performance of global expectancies. *Psychomusicology*, 9 (2), S. 122- 147.
- Schmuckler, M.A. (1997). Expectancy Effects in Memory for Melodies. *Canadian Journal of Experimental Psychology*, 51 (4), S. 292-305.
- Schmuckler, M.A. (1999). Testing Models of Melodic Contour Similarity. *Music Perception*, 16 (3), S. 109-150.
- Schmuckler, M.A. & Boltz, M. (1994). Harmonic and rhythmic influences on musical expectancy. *Perception & Psychophysics*, 56 (3), S. 313-325.
- Schulze, H.-H. (1989). Categorical perception of rhythmic patterns. *Psychological Research*, 51, S. 10-15.

- Selfridge-Field, E. (1998). Conceptual and Representational Issues in Melodic Comparison. In: Hewlett, W.B. & Selfridge-Field, E (Hg.), *Melodic Similarity. Concepts, Procedures, and Applications. Computing in Musicology 11*. Cambridge (MA): MIT Press. 3-64.
- Selfridge-Field, E. (2002). Why is the concept of melodic similarity so important in art music of the European tradition? In: Fingerhut, M. (Hg.), *ISMIR 2002 Conference Proceedings. Third International Conference on Music Information Retrieval*. Paris: IRCAM, S. 307.
- Semal, C. & Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *Journal of the Acoustical Society of America*, 89, S. 2404-2410.
- Semal, C. & Demany, L. (1993). Further evidence for an autonomous processing of pitch in auditory short-term memory. *Journal of the Acoustical Society of America*, 94, S. 1315-1322.
- Serafine, M.L., Crowder, R.G. & Repp, B.H. (1984). Integration of melody and text in memory for songs. *Cognition*, 16, S. 285-303.
- Serafine, M.L., Davidson, J. & Crowder, R.G. (1986). *On the Nature of Melody-Text Integration in Memory for Songs*. *Journal of Memory and Language*, 25, S. 123-135.
- Sergeant, D. (1969). Experimental investigation of absolute pitch. *Journal of Research in Music Education*, 17, S. 135-143.
- Shuter-Dyson, R. (1999). Musical Ability. In: Deutsch, D. (Hg.). *The Psychology of Music, 2nd Edition*. San Diego: Academic Press, S. 627-651.
- Simon, H.A. & Sumner, R.K. (1968). Pattern in music. In: Kleinmuntz, B. (Hg.), *Formal representation of human judgement*. New York: Wiley. S. 219-250.
- Slamecka, N. & Fevreiski, J. (1983). The generation effect when generation fails. *Journal of Verbal Learning and Verbal Behaviour*, 22, S. 153-163.
- Slamecka, N. & Graf, P. (1978). The generation effect: delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4, S. 592-604.
- Sloboda, J.A. (1999). *The Musical Mind. The Cognitive Psychology of Music, 1999 Reprint*. Oxford: University Press.
- Sloboda, J.A. & Parker, D.H.H. (1985). Immediate recall of melodies. In: Howell, P., Cross, I. & West, R. (Hg.), *Musical structure and cognition*. London: Academic Press, S. 143-167.
- Smith, J. (1983). Reproduction and Representation of musical Rhythms: The Effects of musical Skill. In: Rogers, D. & Sloboda, J.A. (Hg.), *The Acquisition of Symbolic Skills*. New York: Plenum, S. 273- 282.
- Smith, K.C. & Cuddy, L.L. (1989). Effects of Metric and Harmonic Rhythm on the Detection of Pitch Alterations in Melodic Sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 15 (3), S. 457-471.
- Snyder, B. (2000). *Music and Memory. An Introduction*. Cambridge (MA): MIT Press.
- Stacho, L. & Nemeth, D. (2001). The Role of Short-Term Memory in Music Skill Acquisition. In: Lappalainen, H (Hg.), *Proceedings of the VIIth International Symposium on Systematic and Comparative Musicology/IIIrd International*

Conference on Cognitive Musicology. Jyväskylä: University of Jyväskylä. S. 171-175.

Steinbeck, W. (1982). Struktur und Ähnlichkeit. Methoden automatisierter Melodieanalyse. *Kieler Schriften zur Musikwissenschaft XXV*. Kassel: Bärenreiter.

Steinke, W.R., Cuddy, L.L. & Holden, R.R. (1997). Dissociation of Musical Tonality and Pitch Memory from Nonmusical Cognitive Abilities. *Canadian Journal of Experimental Psychology*, 51 (4), S. 316-334.

Stephen, G.A. (1994). *String Searching Algorithms*. Singapore: World Scientific.

Stockhausen, K. (1989). *Stockhausen on Music*. Compiled by R. Maconie. New York: Marion Boyars.

Stoffer, Th. (1985). Modelle der kognitiven Repräsentation musikalischer Strukturen. In: Neumann, O. (Hg.), *Perspektiven der Kognitionspsychologie*. Berlin: Springer, S. 147-184.

Stumpf, C. (1883). *Tonpsychologie* (Bd. 1). Leipzig, S. Hirzel.

Sundberg, J. & Lindblom, B. (1976). Generative theories in language and music descriptions. *Cognition*, 4, S. 99-102.

Suppan, W. & Stief, W. (1976). *Melodietypen des deutschen Volksliedes herausgegeben im Auftrag des Deutschen Volksliedarchivs. Band 1*. Tutzing: Schneider.

Swets, J.A. (1973). The relative operating characteristic in psychology. *Science*, 182, S. 990-1000.

Takeuchi, A.H. & Hulse, S.H. (1993). Absolute pitch. *Psychological Bulletin*, 113, S. 345-361.

Taneda, N. & Taneda, R. (1993). *Erziehung zum absoluten Gehör*. Mainz: Schott.

Taylor, J.A. & Pembroke, R.G. (1984). Strategies in Memory for Short Melodies: An Extension of Otto Ortmann's 1933 Study. *Psychomusicology*, 3 (1), S. 16-35.

Temperley, D. (1997). An Algorithm for Harmonic Analysis. *Music Perception*, 15 (1), S. 31-68.

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. Cambridge (MA): MIT Press.

Terhardt, E. (1998). *Akustische Kommunikation*. Berlin: Springer.

Therneau, T.M. & Atkinson, E.J. (1997). *An Introduction to recursive Partitioning Using the RPART Routines*. Internet: <http://www.mayo.edu/hsr/techrpt.html>, Juni 2004.

Thomassen, J.M. (1982). Melodic accent: Experiments and a tentative model. *Journal of the Acoustical Society of America*, 71, S. 1596-1605.

Thompson, B. (1995). Stepwise regression and stepwise discrimination analysis need not apply here. A guidelines tutorial. *Educational and psychological measurement*, 55, S. 525-534.

Tulving, E. (1972). Episodic and semantic memory. In: Tulving, E. & Donaldson, W. (Hg.), *Organization of memory*. New York: Academic Press.

Tulving, E. (1983). *Elements of episodic memory*. Oxford: University Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, 26, S. 1-12.

- Tunks, Th.W., Bowers, D.R. & Eagle, Ch.T. (1993). The effect of stimulus tempo on melodic error detection. *Psychomusicology*, 12, S. 41-51.
- Uitdenbogerd, A.L. (2002). *Music Information Retrieval Technology*. Unveröffentlichte PhD thesis, RMIT University Melbourne Victoria, Australia. Internet: <http://goanna.cs.rmit.edu.au/~alu/research/mythesis.pdf>, März 2003.
- Uitdenbogerd, A.L. & Zobel, J. (2002). Music Ranking Techniques Evaluated. In: Oudshoorn, M. (Hg.), *25th Australasian Computer Science Conference (ACS2002). Conferences in research and Practice in Information Technology, Vol. 4*. Internet: <http://goanna.cs.rmit.edu.au/~jz/fulltext/acsc02uz.pdf>, April, 2003.
- Unyk, A.M. & Carlsen, J.C. (1987). The influence of expectancy on melodic perception. *Psychomusicology*, 7, S. 3-23.
- Vaterrödt-Plünnecke, B. (2001). Episodisches Gedächtnis. In: Pethes, N & Ruchatz, J. (Hg.), *Gedächtnis und Erinnerung. Ein interdisziplinäres Lexikon*. Reinbek: Rowohlt, S. 142-143.
- Važan, P. (2000). Memory for music. An overview. *Systematische Musikwissenschaft*, 7 (1-2), S. 7-31.
- Vispoel, W.P. (1993). The Development and Evaluation of a Computerized Adaptive Test of Tonal Memory. *Journal of Research in Music Education*, 41 (2), S. 111-136.
- Vitouch, O. (2004). Absolutes Gehör. In: Th. H. Stoffer & R. Oerter (Hrsg.), *Allgemeine Musikpsychologie* (Enzyklopädie der Psychologie, Bd. D/VII/1, S. 717-766). Göttingen: Hogrefe.
- Vitouch, O. & Gaugusch, A. (2000). Absolute recognition of musical keys in non-absolute-pitch-possessors. In: Woods, C, Luck, G., Brochard, R., Seddon, F. & Sloboda, J.A. (Hg.), *Proceedings of the 6th International Conference on Music Perception and Cognition*, CD-ROM. Keele: University of Keele.
- Vos, P.G. (1977). Temporal duration factors in the perception of auditory rhythmic patterns. *Scientific Aesthetics*, 1, S. 183-199.
- Wallace, W.P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin*, 63, S. 410-424.
- Wallace, W.T. (1994). Memory for Music: Effect of Melody on Recall of Text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20 (6), S. 1471-1485.
- Ward, W.D. (1999). Absolute pitch. In: Deutsch, D. (Hg.), *The psychology of music*, 2nd Edition. San Diego: Academic Press, S. 265-298.
- Ward, W.D. & Burns, E.M. (1982). Absolute pitch. In: Deutsch, D. (Hg.), *The psychology of music*. New York: Academic Press, S. 431-451.
- Watkins, A.J. & Dyson, M.C. (1985). On the Perceptual Organisation of Tone Sequences and Melodies. In: Howell, P., Cross, I. & West, R. (Hg.), *Musical structure and cognition*. London: Academic Press, S. 71-119.
- Weinert, L. (1929). Untersuchungen über das absolute Gehör. *Archiv für die gesamte Psychologie*, 73, S. 1-128.
- Welker, R.L. (1982). Abstraction of themes from melodic variations. *Journal of Experimental Psychology: Human Perception and Performance*, 8 (3), S. 435-447.
- Werner, H. (1926). Über Mikromelodik und Mikroharmonik. *Zeitschrift für Psychologie*, 98, S. 74-89.

- White, B.W. (1960). Recognition of distorted melodies. *The American Journal of Psychology*, 73, S. 100-107,
- Winslow, R.W. (1949). The psychology of musical memory. *Music-Educators-Journal*, 35 (3), S. 15-16.
- Wiora, W. (1941). Systematik des Umsingens. *Jahrbuch für Volksliedforschung*, 7, S. 128-195.
- Wiora, W. (1977). „Gattungen des Volksliedes“ und „Gattungen der Musik“. Baumann, M.P., Brandl, R.M. & Reinhard, K. (Hg.), *Festschrift für Felix Hoerburger*. Laaber: Laaber, S. 37-44.
- Wundt, W. (1920). *Grundriß der Psychologie*, 14. Auflage. [1. Aufl. 1896] Stuttgart: Kröner.
- Yee, W., Holleran, S. & Jones, M.R.. (1994). Sensivity to event timing in regular and irregular sequences: Influences of musical skill. *Perception & Psychophysics*, 56 (4), S. 461-471.
- Yeston, M. (1976). *The stratification of musical rhythm*. New Haven (CT): Yale University Press.
- Zatorre, R.J. (200). Absolute pitch. a model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience*, 6 (7), S. 692-695.
- Zielinska, H. & Miklaszewski, K. (1992). Memorising Two Melodies of different Style. *Psychology of Music*, 20, S. 95-111.
- Zwicker, E. & Fastl, H. (1999). *Psychoacoustics. Facts and Models*. 2nd Edition. Berlin: Springer.

8 Anhang

A-4.3.1.2: Die 14 verwendeten Melodien der Experimente 1-3

Nr.	ident. Var.	Titel	Interpret	Komponist	Jahr	Quelle
A	/	Goodby My Love, Goodybye	Demis Roussos	Mario Panas	1973	<i>Hits der 70er</i> , KDM-Verlag, 2000
B	1	As Long As You Love me	Backstreet Boys	Martin Sandberg	1997	<i>Hits der 80er und 90er</i> , KDM-Verlag, 1999
C	6	Augen zu und durch	Wolfgang Petry	Petry/Valance/Ackermann	1997	<i>Hits der 80er und 90er</i> , KDM-Verlag, 1999
D	2	Wonderland	Passion Fruit		2000	eigene Transkription
E	5	Danse avec moi	Kosmonova		2000	eigene Transkription
F	3	Bumpin' on sunset	Wes Montgomery	Wes Montgomery	1966	Transkription Pogoda
G	4	Summer is calling	Aquagen		2002	eigene Transkription
H	4	From me to you	The Beatles	Paul McCartney	1963	<i>The News Beatles Complete</i> , Wise Publications, 1992
I	3	Aber Dich gibt's nur einmal für mich	Die Nilsen Brothers	Pit	1965	<i>100 Hits in C-Dur</i> , Musikverlag Monika Hildner
J	5	Die Hände zum Himmel	Die Kolibris	W. van Nimwegen	1998	<i>100 Hits in C-Dur</i> , Musikverlag Monika Hildner
k	/	Swanee River	Bing Crosby	Traditional		<i>Hits und Songs</i> , Edition Metropol Köln, 1982
L	/	Sailor	/	russ. Traditional		aus Sloboda & Parker (1985), nach O'Toole, 1974)
M	2	Du	Peter Maffay	Peter Orloff	1969	<i>Peter Maffay: Heute vor 30 Jahren</i> , Bosworth Edition, 2001
N	5	I want it that way	Backstreet Boys	Andreas Carlson	1999	<i>Hits der 80er und 90er</i> , KDM-Verlag, 1999

A-4.3.1.2: Fragebogen zu den Experimenten 1-3 (nur Angaben zur Person, d.h. ohne Ratingskalen)

Angaben zur Person

1. **Schlüsselziffer** (Die ersten 4 Buchstaben Deines Nachname und die zwei ersten Stellen Deines Geburtsdatums): _____

(z.B. Arnold Schönberg, geb. 13.09.1874 => Scho13)

2. **Alter** _____

3. **Geschlecht:** weiblich ☐ männlich ☐

4. **Hast Du jemals Gehörbildungsunterricht gehabt oder selbst Gehörbildung betrieben?** Ja ☐ nein ☐

a) Falls ja, über welchen Zeitraum hast Du Gehörbildung betrieben? _____ Monate

b) In der aktivsten Phase des Hörtrainings, wie viele Stunden pro Woche hast Du Gehörbildung betrieben? _____ Stunden / Woche

c) Wie viele Stunden pro Woche betreibst Du im Moment Gehörbildung? _____ Stunden / Woche

5. **Wie viele Jahre lang hast Du bisher ein Instrument gespielt bzw. Musik gemacht?** _____ Jahre

6. **Wie viele Stunden pro Woche übst Du im Moment?** _____ Stunden / Woche

7. **Wie viele Stunden pro Woche machst Du im Moment Musik?** (Üben+Proben+Auftritte+Unterricht+Spielen+etc.) _____ Stunden / Woche

8. **In Deiner musikalisch aktivsten Phase, wie viele Stunden pro Woche hast Du geübt?** _____ Stunden / Woche

9. **In Deiner musikalisch aktivsten Phase, wie viele Stunden pro Woche hast Du Musik gemacht?** (Üben+Proben+Auftritte+Unterricht+Spielen+etc.) _____ Stunden / Woche

10. **Wie viele Monate hast Du bisher bezahlten Instrumental-/Gesangsunterricht gehabt?** _____ Monate

11. **Wie viele musikalische Auftritte hast Du bisher absolviert, für die Du bezahlt worden bist?** _____ Auftritte

12. **Wie viele musikalische Solo-Auftritte hast Du bisher absolviert?** _____ Auftritte

13. **Wie viele Auftritte als Begleiter hast du bisher absolviert?** _____ Auftritte

14. **Wie viele Auftritte in einer Gruppe oder Ensemble hast Du bisher absolviert?** _____ Auftritte

15. **Wie viele Stunden pro Woche übst Du Vom-Blatt-Spiel im Moment?** _____ Stunden / Woche

16. In Deiner aktivsten Phase, wie viele Stunden pro Woche hast Du Vom-Blatt-Spiel geübt? _____ Stunden / Woche

A-4.3.1.3.4 Modellierung der Ähnlichkeitsurteile der Experten anhand der Variablen des Versuchsdesigns

1. Modellwerte des Modells mit Fehlerposition (Melodiehälften):

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,904 ^a	,818	,796	,71665

a. Einflußvariablen : (Konstante), FMODT510, FRHYT14, FREIT14, FKONT510, FINTT14, FMODT14, FRHYT510, FINTT510, FKONT14

ANOVA^b

Modell	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1 Regression	170,602	9	18,956	36,909	,000 ^a
Residuen	38,005	74	,514		
Gesamt	208,607	83			

a. Einflußvariablen : (Konstante), FMODT510, FRHYT14, FREIT14, FKONT510, FINTT14, FMODT14, FRHYT510, FINTT510, FKONT14

b. Abhängige Variable: ÄHN_MEAN

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	6,190	,161		38,559	,000
	FRHYT14	-,137	,043	-,201	-3,193	,002
	FINTT14	-,188	,049	-,242	-3,835	,000
	FKONT14	-,269	,066	-,271	-4,100	,000
	FREIT14	-,862	,090	-,494	-9,615	,000
	FMODT14	-,964	,170	-,324	-5,664	,000
	FRHYT510	-,126	,058	-,137	-2,178	,033
	FINTT510	-,411	,057	-,457	-7,150	,000
	FKONT510	-,271	,066	-,264	-4,118	,000
	FMODT510	-,545	,201	-,161	-2,709	,008

a. Abhängige Variable: ÄHN_MEAN

2. Modellwerte des Modells mit der mittleren Fehlerposition der einzelnen Fehlervariablen:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,907 ^a	,823	,799	,71160

a. Einflußvariablen : (Konstante), FPOSMOD, FPOSKON, FPOSREI, FPOSRHY, FINTERVA, FRHYTHM, FPOSINT, FREIHENF, FMODULAT, FKONTUR

ANOVA^b

Modell	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1 Regression	171,642	10	17,164	33,896	,000 ^a
Residuen	36,965	73	,506		
Gesamt	208,607	83			

a. Einflußvariablen : (Konstante), FPOSMOD, FPOSKON, FPOSREI, FPOSRHY, FINTERVA, FRHYTHM, FPOSINT, FREIHENF, FMODULAT, FKONTUR

b. Abhängige Variable: ÄHN_MEAN

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	6,176	,161		38,274	,000
	FRHYTHM	-,123	,037	-,274	-3,336	,001
	FINTERVA	-,206	,045	-,441	-4,590	,000
	FKONTUR	-,226	,064	-,401	-3,563	,001
	FREIHENF	-,690	,169	-,396	-4,075	,000
	FMODULAT	-,530	,184	-,289	-2,875	,005
	FPOSRHY	,019	,058	,025	,319	,751
	FPOSINT	-,158	,068	-,217	-2,324	,023
	FPOSKON	-,080	,094	-,092	-,855	,396
	FPOSREI	-,170	,164	-,099	-1,035	,304
	FPOSMOD	-,118	,087	-,139	-1,354	,180

a. Abhängige Variable: ÄHN_MEAN

3. Modellwerte des Modells mit an Fehlerposition gewichteten Intervallfehlern sowie den vier übrigen Fehlervariablen:

ANOVA^b

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	169,435	6	28,239	55,509	,000 ^a
	Residuen	39,172	77	,509		
	Gesamt	208,607	83			

a. Einflußvariablen : (Konstante), FPOSINT, FRHYTHM, FREIHENF, FMODULAT, FKONTUR, FINTERVA

b. Abhängige Variable: ÄHN_MEAN

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	6,191	,160		38,688	,000
	FRHYTHM	-,125	,023	-,280	-5,484	,000
	FINTERVA	-,202	,042	-,432	-4,759	,000
	FKONTUR	-,274	,030	-,484	-8,984	,000
	FREIHENF	-,857	,088	-,491	-9,698	,000
	FMODULAT	-,746	,092	-,406	-8,087	,000
	FPOSINT	-,157	,066	-,216	-2,357	,021

a. Abhängige Variable: ÄHN_MEAN

4. Modellwerte des Modells mit an Fehlerposition und Fehlerdichte gewichteten Intervallfehlern sowie den vier übrigen (ungewichteten) Fehlervariablen:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,907 ^a	,822	,808	,69455

a. Einflußvariablen : (Konstante), FPOSINTW, FRHYTHM, FREIHENF, FMODULAT, FKONTUR, FINTERVA

ANOVA^b

Modell	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1 Regression	171,462	6	28,577	59,239	,000 ^a
Residuen	37,145	77	,482		
Gesamt	208,607	83			

a. Einflußvariablen : (Konstante), FPOSINTW, FRHYTHM, FREIHENF, FMODULAT, FKONTUR, FINTERVA

b. Abhängige Variable: ÄHN_MEAN

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	6,230	,157		39,689	,000
	FRHYTHM	-,128	,022	-,286	-5,785	,000
	FINTERVA	-,222	,031	-,475	-7,178	,000
	FKONTUR	-,278	,030	-,492	-9,351	,000
	FREIHENF	-,867	,086	-,497	-10,083	,000
	FMODULAT	-,744	,090	-,405	-8,284	,000
	FPOSINTW	-,147	,046	-,210	-3,172	,002

a. Abhängige Variable: ÄHN_MEAN

A-4.4.2 Kontrollexperiment

A-4.4.2.1 Die richtigen Varianten

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,853 ^a	,727	,702	,66111
2	,960 ^b	,922	,906	,37139

a. Einflußvariablen : (Konstante), NGRSUMCO

b. Einflußvariablen : (Konstante), NGRSUMCO, HARMCORE

ANOVA^c

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	12,796	1	12,796	29,277	,000 ^a
	Residuen	4,808	11	,437		
	Gesamt	17,604	12			
2	Regression	16,224	2	8,112	58,814	,000 ^b
	Residuen	1,379	10	,138		
	Gesamt	17,604	12			

a. Einflußvariablen : (Konstante), NGRSUMCO

b. Einflußvariablen : (Konstante), NGRSUMCO, HARMCORE

c. Abhängige Variable: KÄHNMEAN

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	2,970	,332		8,944	,000
	NGRSUMCO	6,424	1,187	,853	5,411	,000
2	(Konstante)	2,294	,231		9,944	,000
	NGRSUMCO	4,250	,797	,564	5,333	,000
	HARMCORE	2,154	,432	,527	4,986	,001

a. Abhängige Variable: KÄHNMEAN

A-4.4.2.2 Richtige und falsche Varianten

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,917 ^a	,841	,826	,68022

a. Einflußvariablen : (Konstante), NGRCOORD, RAWEDW

ANOVA^b

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	51,344	2	25,672	55,484	,000 ^a
	Residuen	9,717	21	,463		
	Gesamt	61,060	23			

a. Einflußvariablen : (Konstante), NGRCOORD, RAWEDW

b. Abhängige Variable: KÄHNMEAN

Koeffizienten^a

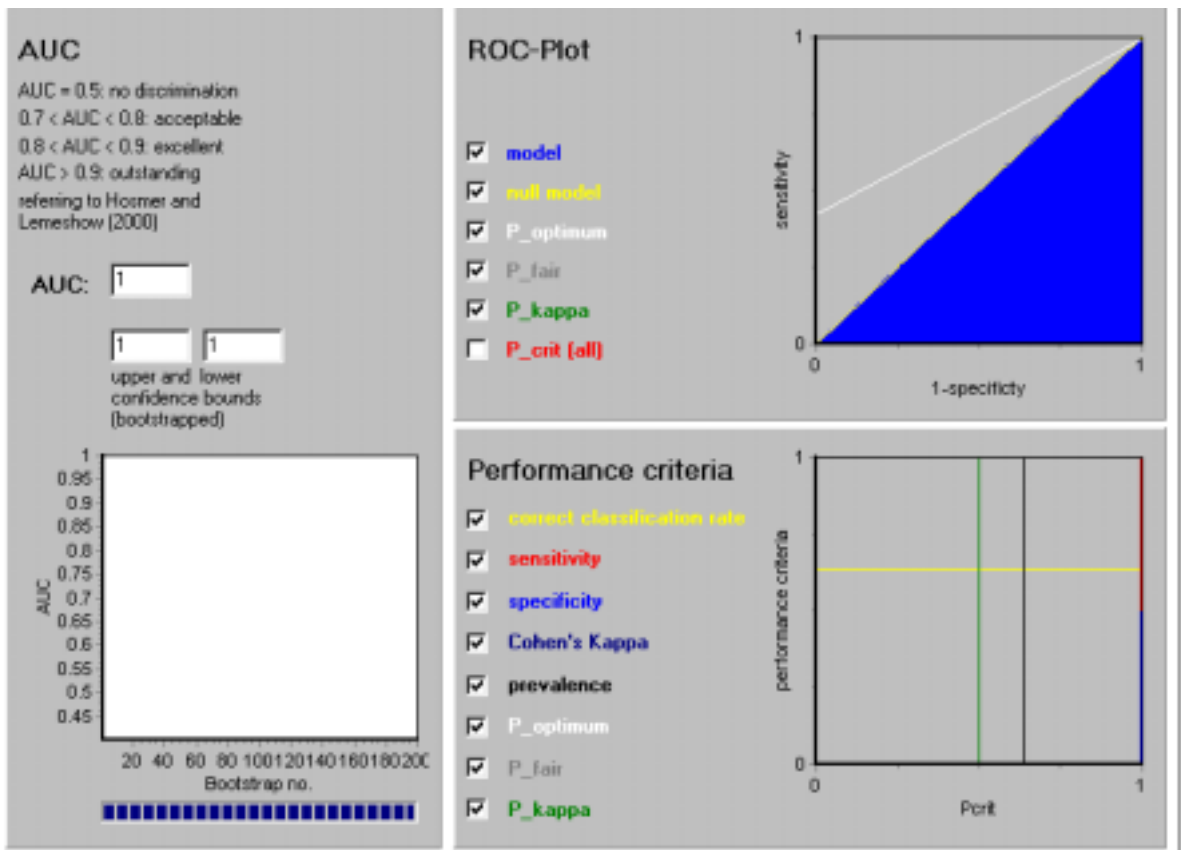
Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	1,136	,295		3,855	,001
	RAWEDW	3,141	,886	,458	3,544	,002
	NGRCOORD	2,777	,683	,525	4,064	,001

a. Abhängige Variable: KÄHNMEAN

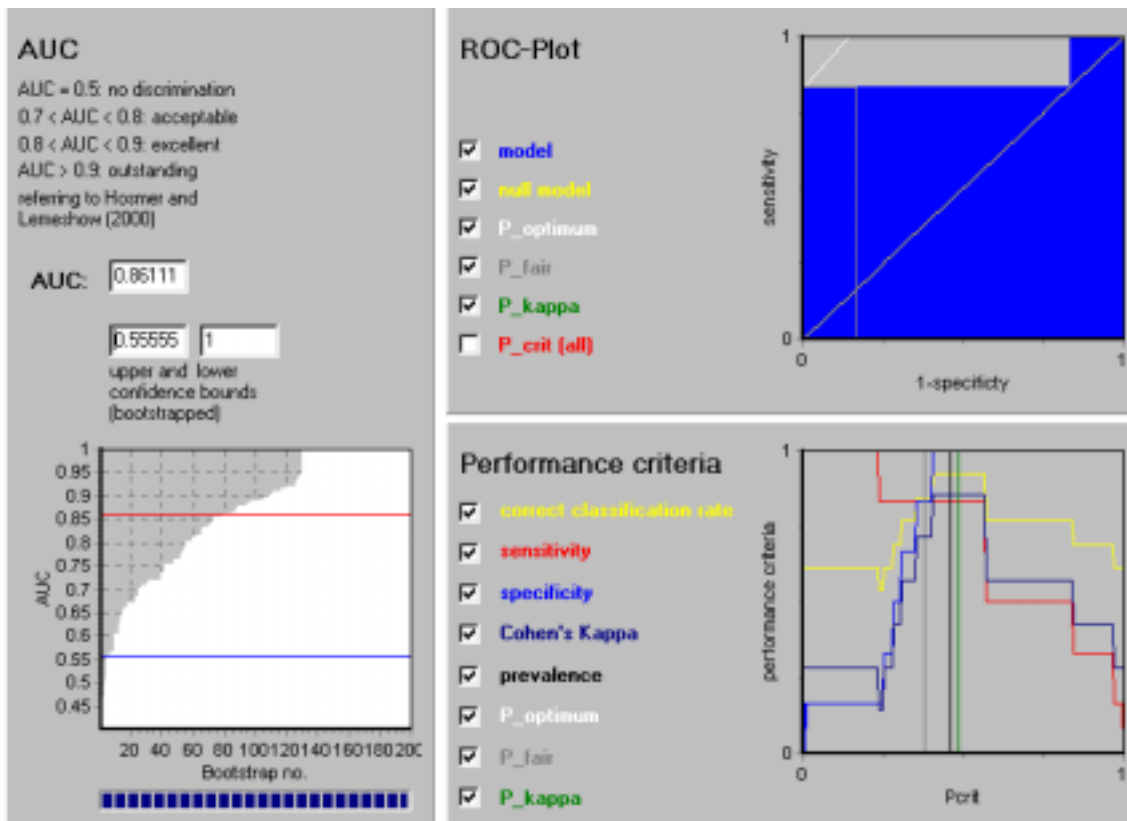
A-4.4.2.2 Plots zur Diskriminationsleistung der Ähnlichkeitsmaße

Die Plots sind mit Hilfe des frei verfügbaren Programms *ROC_AUC* (von Boris Schröder vom Institut für Geoökologie der Universität Potsdam) erstellt worden.

Zunächst wird der perfekte Diskriminationszusammenhang beim Maß *opti2* dargestellt, bei dem die *area* den gesamten ROC-Raum ausmacht (weißes Quadrat links unten in der Abbildung). Der zugehörige AUC-Wert ist das Maximum 1. Die exakt selben Grafiken ergaben sich natürlich auch für die Urteile der Musikexperten, die gleich gut diskriminierten.



Die analoge Grafik ist hier für die Ähnlichkeitswerte von *rawed* hinsichtlich der Varianten von Melodie C wiedergegeben. Es lässt sich deutlich die geringere *area* (weißer Bereich) erkennen. *Rawed* überschätzt die Ähnlichkeit von ‚falschen‘ Varianten, wie deutlich an den Plots abzulesen ist.



A-4.4.3 Experiment 3

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,820 ^a	,672	,661	1,24942
2	,896 ^b	,803	,790	,98402

a. Einflußvariablen : (Konstante), NGRCOORF

b. Einflußvariablen : (Konstante), NGRCOORF, RAWED

ANOVA

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	95,844	1	95,844	61,397	,000 ^a
	Residuen	46,831	30	1,561		
	Gesamt	142,675	31			
2	Regression	114,594	2	57,297	59,173	,000 ^b
	Residuen	28,081	29	,968		
	Gesamt	142,675	31			

a. Einflußvariablen : (Konstante), NGRCOORF

b. Einflußvariablen : (Konstante), NGRCOORF, RAWED

c. Abhängige Variable: VPNMEAN3

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	2,246	,477		4,705	,000
	NGRCOORF	6,387	,815	,820	7,836	,000
2	(Konstante)	,422	,560		,753	,457
	NGRCOORF	4,048	,833	,519	4,857	,000
	RAWED	4,420	1,004	,471	4,400	,000

a. Abhängige Variable: VPNMEAN3

A-5.4.7: Kreuztabelle zu Melodienummer und richtigen Antworten

MELODYNO * RATECORR Kreuztabelle

			RATECORR		Gesamt
			,00	1,00	
MELODYNO	1	Anzahl	12	16	28
		% von MELODYNO	42,9%	57,1%	100,0%
	2	Anzahl	8	19	27
		% von MELODYNO	29,6%	70,4%	100,0%
	3	Anzahl	19	9	28
		% von MELODYNO	67,9%	32,1%	100,0%
	4	Anzahl	3	25	28
		% von MELODYNO	10,7%	89,3%	100,0%
	5	Anzahl	4	24	28
		% von MELODYNO	14,3%	85,7%	100,0%
	6	Anzahl	10	18	28
		% von MELODYNO	35,7%	64,3%	100,0%
	7	Anzahl	7	19	26
		% von MELODYNO	26,9%	73,1%	100,0%
	8	Anzahl	9	16	25
		% von MELODYNO	36,0%	64,0%	100,0%
	9	Anzahl	4	24	28
		% von MELODYNO	14,3%	85,7%	100,0%
	10	Anzahl	11	17	28
		% von MELODYNO	39,3%	60,7%	100,0%
	11	Anzahl	11	17	28
		% von MELODYNO	39,3%	60,7%	100,0%
	12	Anzahl	2	26	28
		% von MELODYNO	7,1%	92,9%	100,0%
	13	Anzahl	10	18	28
		% von MELODYNO	35,7%	64,3%	100,0%
	14	Anzahl	12	16	28
		% von MELODYNO	42,9%	57,1%	100,0%
	15	Anzahl	1	27	28
		% von MELODYNO	3,6%	96,4%	100,0%
	16	Anzahl	3	25	28
		% von MELODYNO	10,7%	89,3%	100,0%
	17	Anzahl	13	15	28
		% von MELODYNO	46,4%	53,6%	100,0%
	18	Anzahl	6	22	28
		% von MELODYNO	21,4%	78,6%	100,0%
	19	Anzahl	5	23	28
		% von MELODYNO	17,9%	82,1%	100,0%
	20	Anzahl	8	20	28
		% von MELODYNO	28,6%	71,4%	100,0%
	21	Anzahl	10	18	28
		% von MELODYNO	35,7%	64,3%	100,0%
	22	Anzahl	5	23	28
		% von MELODYNO	17,9%	82,1%	100,0%
	23	Anzahl	6	22	28
		% von MELODYNO	21,4%	78,6%	100,0%
	24	Anzahl	10	18	28
		% von MELODYNO	35,7%	64,3%	100,0%
	25	Anzahl	19	7	26
		% von MELODYNO	73,1%	26,9%	100,0%
	26	Anzahl	5	22	27
		% von MELODYNO	18,5%	81,5%	100,0%
	27	Anzahl	15	13	28
		% von MELODYNO	53,6%	46,4%	100,0%
	28	Anzahl	14	13	27
		% von MELODYNO	51,9%	48,1%	100,0%
	29	Anzahl	5	21	26
		% von MELODYNO	19,2%	80,8%	100,0%
	30	Anzahl	5	25	30
		% von MELODYNO	16,7%	83,3%	100,0%
	31	Anzahl	18	10	28
		% von MELODYNO	64,3%	35,7%	100,0%
	32	Anzahl	13	15	28
		% von MELODYNO	46,4%	53,6%	100,0%
	33	Anzahl	10	17	27
		% von MELODYNO	37,0%	63,0%	100,0%
	34	Anzahl	6	22	28
		% von MELODYNO	21,4%	78,6%	100,0%
	35	Anzahl	9	19	28
		% von MELODYNO	32,1%	67,9%	100,0%
	36	Anzahl	10	18	28
		% von MELODYNO	35,7%	64,3%	100,0%
Gesamt		Anzahl	318	679	997
		% von MELODYNO	31,9%	68,1%	100,0%

A-5.5.2 Vierfaktorielles Modell ohne Datenaggregation: Geschätzte Randmittel und Scheffé-Tests

Schätzungen

Abhängige Variable: complete rating data

MUSICTACT	Mittelwert	Standardfehler	95% Konfidenzintervall	
			Untergrenze	Obergrenze
0	3,300 ^a	,080	3,143	3,456
1	3,145 ^a	,065	3,018	3,273
2	3,004 ^a	,079	2,849	3,160

a. Basiert auf Randmittel der geänderten Grundgesamtheit.

Mehrfachvergleiche

Abhängige Variable: complete rating data

Scheffé

(I) MUSICTACT	(J) MUSICTACT	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95% Konfidenzintervall	
					Untergrenze	Obergrenze
0	1	,13	,095	,369	-,10	,37
	2	,19	,104	,193	-,07	,44
1	0	-,13	,095	,369	-,37	,10
	2	,05	,095	,849	-,18	,29
2	0	-,19	,104	,193	-,44	,07
	1	-,05	,095	,849	-,29	,18

Basiert auf beobachteten Mittelwerten.

Schätzungen

Abhängige Variable: complete rating data

Manipulation class	Mittelwert	Standardfehler	95% Konfidenzintervall	
			Untergrenze	Obergrenze
none	4,028 ^a	,069	3,893	4,163
nonac	3,085 ^a	,070	2,948	3,223
accent	2,922 ^a	,069	2,786	3,057

a. Basiert auf Randmittel der geänderten Grundgesamtheit.

Mehrfachvergleiche

Abhängige Variable: complete rating data
Scheffé

(I) Manipulation class	(J) Manipulation class	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95% Konfidenzintervall	
					Untergrenze	Obergrenze
none	nonac	,92*	,096	,000	,68	1,15
	accent	1,09*	,096	,000	,85	1,32
nonac	none	-,92*	,096	,000	-1,15	-,68
	accent	,17	,096	,201	-,06	,41
accent	none	-1,09*	,096	,000	-1,32	-,85
	nonac	-,17	,096	,201	-,41	,06

Basiert auf beobachteten Mittelwerten.

*. Die mittlere Differenz ist auf der Stufe ,05 signifikant.

Schätzungen

Abhängige Variable: complete rating data

Manipulation type	Mittelwert	Standardfehler	95% Konfidenzintervall	
			Untergrenze	Obergrenze
none	4,028 ^a	,069	3,893	4,163
rhythm	3,222 ^a	,087	3,052	3,391
interval	2,894 ^a	,084	2,728	3,059
contour	2,895 ^a	,085	2,729	3,061

a. Basiert auf Randmittel der geänderten Grundgesamtheit.

Mehrfachvergleiche

Abhängige Variable: complete rating data
Scheffé

(I) Manipulation type	(J) Manipulation type	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95% Konfidenzintervall	
					Untergrenze	Obergrenze
none	rhythm	,77*	,108	,000	,47	1,07
	interval	1,11*	,107	,000	,81	1,41
	contour	1,13*	,107	,000	,83	1,43
rhythm	none	-,77*	,108	,000	-1,07	-,47
	interval	,34*	,118	,038	,01	,68
	contour	,36*	,118	,028	,03	,69
interval	none	-1,11*	,107	,000	-1,41	-,81
	rhythm	-,34*	,118	,038	-,68	-,01
	contour	,01	,118	1,000	-,32	,34
contour	none	-1,13*	,107	,000	-1,43	-,83
	rhythm	-,36*	,118	,028	-,69	-,03
	interval	-,01	,118	1,000	-,34	,32

Basiert auf beobachteten Mittelwerten.

*. Die mittlere Differenz ist auf der Stufe ,05 signifikant.

Schätzungen

Abhängige Variable: complete rating data

Position of comparison melody	Mittelwert	Standardf ehler	95% Konfidenzintervall	
			Untergrenze	Obergrenze
1st part	2,973 ^a	,061	2,854	3,093
2nd part	3,326 ^a	,061	3,206	3,446

a. Basiert auf Randmittel der geänderten Grundgesamtheit.

5. MUSICACT * Position of comparison melody

Abhängige Variable: complete rating data

MUSICACT	Position of comparison melody	Mittelwert	Standardf ehler	95% Konfidenzintervall	
				Untergrenze	Obergrenze
0	1st part	3,045 ^a	,110	2,830	3,260
	2nd part	3,554 ^a	,110	3,339	3,769
1	1st part	2,868 ^a	,090	2,692	3,044
	2nd part	3,423 ^a	,090	3,246	3,600
2	1st part	3,008 ^a	,109	2,793	3,222
	2nd part	3,001 ^a	,109	2,787	3,215

a. Basiert auf Randmittel der geänderten Grundgesamtheit.

6. Manipulation class * MUSICACT

Abhängige Variable: complete rating data

Manipulation class	MUSICACT	Mittelwert	Standardf ehler	95% Konfidenzintervall	
				Untergrenze	Obergrenze
none	0	3,979 ^a	,126	3,731	4,227
	1	3,887 ^a	,104	3,683	4,090
	2	4,219 ^a	,126	3,971	4,467
nonac	0	3,110 ^a	,130	2,856	3,365
	1	3,143 ^a	,106	2,935	3,350
	2	3,002 ^a	,127	2,752	3,252
accent	0	3,263 ^a	,127	3,014	3,511
	1	2,901 ^a	,103	2,698	3,104
	2	2,601 ^a	,127	2,351	2,851

a. Basiert auf Randmittel der geänderten Grundgesamtheit.

7. Manipulation type * MUSICACT

Abhängige Variable: complete rating data

Manipulation type	MUSICACT	Mittelwert	Standardf ehler	95% Konfidenzintervall	
				Untergrenze	Obergrenze
none	0	3,979 ^a	,126	3,731	4,227
	1	3,887 ^a	,104	3,683	4,090
	2	4,219 ^a	,126	3,971	4,467
rhythm	0	3,195 ^a	,160	2,880	3,510
	1	3,300 ^a	,127	3,051	3,548
	2	3,170 ^a	,160	2,856	3,485
interval	0	3,229 ^a	,153	2,930	3,529
	1	2,882 ^a	,129	2,630	3,135
	2	2,569 ^a	,156	2,263	2,876
contour	0	3,135 ^a	,157	2,826	3,444
	1	2,884 ^a	,129	2,631	3,137
	2	2,665 ^a	,152	2,368	2,963

a. Basiert auf Randmittel der geänderten Grundgesamtheit.

8. Manipulation class * Position of comparison melody

Abhängige Variable: complete rating data

Manipulation class	Position of comparison melody	Mittelwert	Standardf ehler	95% Konfidenzintervall	
				Untergrenze	Obergrenze
none	1st part	3,902 ^a	,097	3,711	4,092
	2nd part	4,155 ^a	,097	3,965	4,344
nonac	1st part	2,877 ^a	,098	2,683	3,070
	2nd part	3,294 ^a	,098	3,100	3,487
accent	1st part	2,761 ^a	,097	2,571	2,951
	2nd part	3,082 ^a	,097	2,891	3,273

a. Basiert auf Randmittel der geänderten Grundgesamtheit.

9. Manipulation type * Position of comparison melody

Abhängige Variable: complete rating data

Manipulation type	Position of comparison melody	Mittelwert	Standardf ehler	95% Konfidenzintervall	
				Untergrenze	Obergrenze
none	1st part	3,902 ^a	,097	3,711	4,092
	2nd part	4,155 ^a	,097	3,965	4,344
rhythm	1st part	2,989 ^a	,121	2,751	3,227
	2nd part	3,454 ^a	,121	3,217	3,692
interval	1st part	2,573 ^a	,118	2,341	2,804
	2nd part	3,215 ^a	,120	2,980	3,449
contour	1st part	2,895 ^a	,119	2,661	3,128
	2nd part	2,895 ^a	,119	2,662	3,128

a. Basiert auf Randmittel der geänderten Grundgesamtheit.

10. Manipulation class * Manipulation type

Abhängige Variable: complete rating data

Manipulation class	Manipulation type	Mittelwert	Standardfehler	95% Konfidenzintervall	
				Untergrenze	Obergrenze
none	none	4,028	,069	3,893	4,163
	rhythm	. ^a	.	.	.
	interval	. ^a	.	.	.
	contour	. ^a	.	.	.
nonac	none	. ^a	.	.	.
	rhythm	3,240	,124	2,997	3,484
	interval	3,246	,119	3,012	3,479
	contour	2,769	,119	2,536	3,002
accent	none	. ^a	.	.	.
	rhythm	3,203	,119	2,970	3,436
	interval	2,542	,119	2,309	2,775
	contour	3,020	,119	2,786	3,254

a. Diese Stufenkombination der Faktoren wurde nicht beobachtet, daher ist das entsprechende Grundgesamtheits-Randmittel nicht schätzbar.

A-5.6.2.1 Vollständiges lineares Modell als Input zur Variablenselektion

RATDACOM ~ ZMUSICCO+ZMUSICAC+NUMMEEND+DURTEMEL+LENTEMEL+SIMILAR+ NUMNOEND+SECSSEND+POCOMEL+JOINT52+MANRHYTH+MANINTER+MANCONT+ NOTEREST+HARMFUNC+CHANGERA+MANSECEN%in%SIMILAR+MANOTEPH%in%SIMILAR+MANPHMEL%in%SIMILAR+ZMUSICCO*ZMUSICAC+ZMUSICCO*NUMMEEND+ZMUSICCO*DURTEMEL+ZMUSICCO*JOINT52+ ZMUSICCO*NOTEREST+ZMUSICCO*HARMFUNC+ZMUSICCO*CHANGERA+ZMUSICCO*LENTEMEL+ ZMUSICCO* NUMNOEND+ZMUSICCO* SECSSEND+ ZMUSICCO* POCOMEL+ZMUSICCO*SIMILAR+ZMUSICCO*MANRHYTH+ZMUSICCO* MANINTER+ZMUSICCO* MANCONT+ZMUSICAC*NUMMEEND+ZMUSICAC*DURTEMEL+ZMUSICAC*LENTEMEL+ZMUSICAC*JOINT52+ZMUSICAC*NOTEREST+ZMUSICAC*HARMFUNC+ZMUSICAC*CHANGERA+ ZMUSICAC* NUMNOEND+ZMUSICAC* SECSSEND+ ZMUSICAC* POCOMEL+ZMUSICAC* SIMILAR+ ZMUSICAC* MANRHYTH+ZMUSICAC* MANINTER+ ZMUSICAC* MANCONT+NUMMEEND*DURTEMEL+NUMMEEND*LENTEMEL+NUMMEEND*JOINT52+NUMMEEND*NOTEREST+NUMMEEND*HARMFUNC+NUMMEEND*CHANGERA+NUMMEEND* NUMNOEND+ NUMMEEND* SECSSEND+NUMMEEND* POCOMEL+ NUMMEEND* SIMILAR+NUMMEEND* MANRHYTH+ NUMMEEND* MANINTER+NUMMEEND*MANCONT+DURTEMEL*LENTEMEL+DURTEMEL*JOINT52+DURTEMEL*NOTEREST+

DURTEMEL*HARMFUNC+DURTEMEL*CHANGERA+
 DURTEMEL* NUMNOEND+ DURTEMEL* SECSSEND+
 DURTEMEL* POCOMEL+ DURTEMEL* SIMILAR+
 DURTEMEL* MANRHYTH+ DURTEMEL* MANINTER+
 DURTEMEL*MANCONT+LENTEMEL*JOINT52+
 LENTEMEL*NOTEREST+LENTEMEL*HARMFUNC+
 LENTEMEL*CHANGERA+ LENTEMEL* NUMNOEND+
 LENTEMEL* SECSSEND+ LENTEMEL* POCOMEL+
 LENTEMEL* SIMILAR+ LENTEMEL* MANRHYTH+
 LENTEMEL* MANINTER+ LENTEMEL* MANCONT+
 JOINT52*NOTEREST+JOINT52*HARMFUNC+
 JOINT52*CHANGERA+ JOINT52* NUMNOEND+
 JOINT52* SECSSEND+ JOINT52* POCOMEL+
 JOINT52* SIMILAR+ JOINT52* MANRHYTH+
 JOINT52* MANINTER+ JOINT52* MANCONT+
 NOTEREST*HARMFUNC+NOTEREST*CHANGERA+
 NOTEREST* NUMNOEND+ NOTEREST* SECSSEND+
 NOTEREST* POCOMEL+ NOTEREST* SIMILAR+
 NOTEREST* MANRHYTH+ NOTEREST* MANINTER+
 NOTEREST* MANCONT+HARMFUNC*CHANGERA+
 HARMFUNC* NUMNOEND+ HARMFUNC* SECSSEND+
 HARMFUNC* POCOMEL+ HARMFUNC* SIMILAR+
 HARMFUNC* MANRHYTH+ HARMFUNC* MANINTER+
 HARMFUNC* MANCONT+CHANGERA* NUMNOEND+
 CHANGERA* SECSSEND+ CHANGERA* POCOMEL+
 CHANGERA* SIMILAR+ CHANGERA* MANRHYTH+
 CHANGERA* MANINTER+ CHANGERA* MANCONT+
 NUMNOEND*SECSSEND+NUMNOEND*POCOMEL+
 NUMNOEND*SIMILAR+NUMNOEND*MANRHYTH+
 NUMNOEND*MANINTER+NUMNOEND*MANCONT+
 SECSSEND* POCOMEL+ SECSSEND* SIMILAR+
 SECSSEND* MANRHYTH+SECSSEND* MANINTER+
 SECSSEND* MANCONT+ POCOMEL* SIMILAR+
 POCOMEL* MANRHYTH+ POCOMEL* MANINTER+
 POCOMEL* MANCONT+ SIMILAR* MANRHYTH+
 SIMILAR* MANINTER+ SIMILAR* MANCONT+
 MANRHYTH* MANINTER+ MANRHYTH* MANCONT+
 MANINTER*MANCONT

A-5.6.3.2 Klassifizierungsbaum

Baumzusammenfassung:

```
> tree1 class
n= 665
```

```
node), split, n, loss, yval, (yprob)
* denotes terminal node
```

- 1) root 665 477 3 (0.11 0.16 0.28 0.26 0.1 0.074)
- 2) SIMILAR < 0.86035 406 261 3 (0.16 0.21 0.36 0.2 0.052 0.02)

4) ZMUSICCO < 0.338823 288 186 3 (0.2 0.16 0.35 0.22 0.049 0.024)
 8) ZMUSICCO >= 0.1411808 31 14 1 (0.55 0.032 0.16 0.13 0.065 0.065) *
 9) ZMUSICCO < 0.1411808 257 160 3 (0.16 0.18 0.38 0.23 0.047 0.019) *
 5) ZMUSICCO >= 0.338823 118 75 3 (0.051 0.34 0.36 0.18 0.059 0.0085)
 10) ZMUSICCO < 0.7213568 57 31 2 (0.053 0.46 0.21 0.21 0.07 0) *
 11) ZMUSICCO >= 0.7213568 61 30 3 (0.049 0.23 0.51 0.15 0.049 0.016) *
 3) SIMILAR >= 0.86035 259 166 4 (0.042 0.089 0.17 0.36 0.19 0.16)
 6) DURTEMEL < 14.5 181 104 4 (0.039 0.1 0.14 0.43 0.15 0.14) *
 7) DURTEMEL >= 14.5 78 58 5 (0.051 0.051 0.23 0.21 0.26 0.21)
 14) ZMUSICCO >= 0.484234 18 9 5 (0.056 0 0.33 0.11 0.5 0) *
 15) ZMUSICCO < 0.484234 60 44 6 (0.05 0.067 0.2 0.23 0.18 0.27)
 30) NUMNOEND >= 33 16 9 3 (0.13 0.063 0.44 0.25 0.13 0) *
 31) NUMNOEND < 33 44 28 6 (0.023 0.068 0.11 0.23 0.2 0.36) *

Klassifizierungsbaum: Vollständige Beschreibung:

Call:

```

rpart(formula = RATDACOM ~ SIMILAR + MANRHYTH + MANINTER +
MANCONT +
  NUMNOEND + NUMMEEND + SECSSEND + DURTEMEL + LENTEMEL +
POCOMEL +
  JOINT52 + MANSECEN + MANOTEPH + MANPHMEL + NOTEREST +
HARMFUNC +
  CHANGERA + ZMUSICCO + ZMUSICAC, data = treedata, method = "class",
  control = list(maxsurrogate = 18, cp = 0.005))
n= 665
  
```

	CP	nsplit	rel error	xerror	xstd
1	0.104821803	0	1.0000000	1.0000000	0.02434495
2	0.018169113	1	0.8951782	0.9014675	0.02584276
3	0.011180992	4	0.8406709	0.8930818	0.02594027
4	0.008385744	7	0.8071279	0.9014675	0.02584276
5	0.006289308	10	0.7819706	0.8867925	0.02601050
6	0.005241090	14	0.7568134	0.8972746	0.02589207
7	0.005000000	18	0.7358491	0.9035639	0.02581769

Node number 1: 665 observations, complexity param=0.1048218

predicted class=3 expected loss=0.7172932

class counts: 74 109 188 176 69 49

probabilities: 0.111 0.164 0.283 0.265 0.104 0.074

left son=2 (406 obs) right son=3 (259 obs)

Primary splits:

SIMILAR < 0.86035	to the left, improve=19.818720, (0 missing)
JOINT52 < 0.91795	to the left, improve=19.818720, (0 missing)
CHANGERA < 0.5	to the right, improve= 7.419152, (0 missing)
HARMFUNC < 0.5	to the right, improve= 7.098724, (0 missing)
LENTEMEL < 32.5	to the left, improve= 6.466088, (0 missing)

Surrogate splits:

JOINT52 < 0.91795	to the left, agree=1.000, adj=1.000, (0 split)
LENTEMEL < 32.5	to the left, agree=0.726, adj=0.297, (0 split)

CHANGERA < 0.5 to the right, agree=0.689, adj=0.201, (0 split)
 SECSSEND < 30 to the left, agree=0.674, adj=0.162, (0 split)
 HARMFUNC < 0.5 to the right, agree=0.674, adj=0.162, (0 split)
 NOTEREST < 0.5 to the right, agree=0.662, adj=0.131, (0 split)
 DURTEMEL < 15.5 to the left, agree=0.645, adj=0.089, (0 split)
 MANCONT < 0.5 to the right, agree=0.618, adj=0.019, (0 split)

Node number 2: 406 observations, complexity param=0.01816911

predicted class=3 expected loss=0.6428571

class counts: 63 86 145 83 21 8

probabilities: 0.155 0.212 0.357 0.204 0.052 0.020

left son=4 (288 obs) right son=5 (118 obs)

Primary splits:

ZMUSICCO < 0.338823 to the left, improve=4.656156, (0 missing)
 JOINT52 < 0.61805 to the left, improve=3.812898, (0 missing)
 SIMILAR < 0.5337 to the left, improve=3.144317, (0 missing)
 ZMUSICAC < 0.3710032 to the right, improve=3.000817, (0 missing)
 MANSECEN < 8.5875 to the right, improve=2.808501, (0 missing)

Surrogate splits:

ZMUSICAC < 1.132064 to the left, agree=0.749, adj=0.136, (0 split)

Node number 3: 259 observations, complexity param=0.01118099

predicted class=4 expected loss=0.6409266

class counts: 11 23 43 93 48 41

probabilities: 0.042 0.089 0.166 0.359 0.185 0.158

left son=6 (181 obs) right son=7 (78 obs)

Primary splits:

DURTEMEL < 14.5 to the left, improve=4.087507, (0 missing)
 ZMUSICCO < 1.733954 to the left, improve=3.485675, (0 missing)
 SECSSEND < 36.75 to the right, improve=3.200739, (0 missing)
 LENTEMEL < 15.5 to the left, improve=3.200739, (0 missing)
 ZMUSICAC < -1.861554 to the right, improve=3.150095, (0 missing)

Surrogate splits:

LENTEMEL < 28 to the left, agree=0.784, adj=0.282, (0 split)
 NUMNOEND < 58.5 to the left, agree=0.776, adj=0.256, (0 split)
 NUMMEEND < 1.5 to the left, agree=0.776, adj=0.256, (0 split)
 SECSSEND < 39 to the left, agree=0.776, adj=0.256, (0 split)
 SIMILAR < 0.8717 to the right, agree=0.772, adj=0.244, (0 split)
 MANRHYTH < 0.5 to the left, agree=0.707, adj=0.026, (0 split)
 JOINT52 < 0.98835 to the right, agree=0.707, adj=0.026, (0 split)
 NOTEREST < 0.5 to the left, agree=0.707, adj=0.026, (0 split)

Node number 4: 288 observations, complexity param=0.01816911

predicted class=3 expected loss=0.6458333

class counts: 57 46 102 62 14 7

probabilities: 0.198 0.160 0.354 0.215 0.049 0.024

left son=8 (31 obs) right son=9 (257 obs)

Primary splits:

ZMUSICCO < 0.1411808 to the right, improve=6.447130, (0 missing)
 ZMUSICAC < 0.2085403 to the right, improve=3.224413, (0 missing)
 JOINT52 < 0.61805 to the left, improve=2.675534, (0 missing)

SIMILAR < 0.5337 to the left, improve=2.558829, (0 missing)
 LENTEMEL < 27.5 to the left, improve=2.300534, (0 missing)
 Surrogate splits:
 ZMUSICAC < 0.3166205 to the right, agree=0.896, adj=0.032, (0 split)

Node number 5: 118 observations, complexity param=0.01816911
 predicted class=3 expected loss=0.6355932
 class counts: 6 40 43 21 7 1
 probabilities: 0.051 0.339 0.364 0.178 0.059 0.008
 left son=10 (57 obs) right son=11 (61 obs)

Primary splits:
 ZMUSICCO < 0.7213568 to the left, improve=4.262519, (0 missing)
 SIMILAR < 0.4554 to the left, improve=3.868045, (0 missing)
 ZMUSICAC < -0.2985999 to the right, improve=3.765869, (0 missing)
 JOINT52 < 0.66875 to the left, improve=3.417711, (0 missing)
 LENTEMEL < 25.5 to the left, improve=3.056275, (0 missing)

Surrogate splits:
 ZMUSICAC < -0.03573944 to the left, agree=0.627, adj=0.228, (0 split)
 SECSSEND < 26.375 to the right, agree=0.551, adj=0.070, (0 split)
 MANSECEN < 32.75 to the right, agree=0.551, adj=0.070, (0 split)
 MANOTEPH < 4.5 to the left, agree=0.551, adj=0.070, (0 split)
 SIMILAR < 0.41425 to the left, agree=0.542, adj=0.053, (0 split)
 NUMNOEND < 44 to the right, agree=0.534, adj=0.035, (0 split)
 NUMMEEND < 1.5 to the right, agree=0.534, adj=0.035, (0 split)
 DURTEMEL < 16 to the right, agree=0.534, adj=0.035, (0 split)
 MANPHMEL < 1.75 to the right, agree=0.534, adj=0.035, (0 split)
 NOTEREST < 0.5 to the left, agree=0.534, adj=0.035, (0 split)
 LENTEMEL < 15.5 to the left, agree=0.525, adj=0.018, (0 split)
 POCOMEL < 1.5 to the left, agree=0.525, adj=0.018, (0 split)
 JOINT52 < 0.85295 to the right, agree=0.525, adj=0.018, (0 split)

Node number 6: 181 observations, complexity param=0.00524109
 predicted class=4 expected loss=0.5745856
 class counts: 7 19 25 77 28 25
 probabilities: 0.039 0.105 0.138 0.425 0.155 0.138
 left son=12 (13 obs) right son=13 (168 obs)

Primary splits:
 ZMUSICCO < 1.733954 to the right, improve=4.298940, (0 missing)
 ZMUSICAC < 0.3710032 to the left, improve=3.274187, (0 missing)
 DURTEMEL < 9.5 to the left, improve=1.945213, (0 missing)
 SECSSEND < 36.75 to the right, improve=1.822707, (0 missing)
 LENTEMEL < 14 to the left, improve=1.822707, (0 missing)

Node number 7: 78 observations, complexity param=0.01118099
 predicted class=5 expected loss=0.7435897
 class counts: 4 4 18 16 20 16
 probabilities: 0.051 0.051 0.231 0.205 0.256 0.205
 left son=14 (18 obs) right son=15 (60 obs)

Primary splits:
 ZMUSICCO < 0.484234 to the right, improve=2.888034, (0 missing)
 ZMUSICAC < 1.132064 to the right, improve=2.845354, (0 missing)

NUMNOEND < 33 to the right, improve=2.105659, (0 missing)
 NUMMEEND < 1 to the right, improve=2.105659, (0 missing)
 SECSSEND < 20 to the right, improve=2.105659, (0 missing)
 Surrogate splits:
 ZMUSICAC < 1.132064 to the right, agree=0.833, adj=0.278, (0 split)

Node number 8: 31 observations
 predicted class=1 expected loss=0.4516129
 class counts: 17 1 5 4 2 2
 probabilities: 0.548 0.032 0.161 0.129 0.065 0.065

Node number 9: 257 observations, complexity param=0.006289308
 predicted class=3 expected loss=0.6225681
 class counts: 40 45 97 58 12 5
 probabilities: 0.156 0.175 0.377 0.226 0.047 0.019
 left son=18 (72 obs) right son=19 (185 obs)
 Primary splits:
 JOINT52 < 0.61805 to the left, improve=4.010088, (0 missing)
 ZMUSICCO < -0.2750667 to the left, improve=2.795244, (0 missing)
 LENTEMEL < 27.5 to the left, improve=2.700109, (0 missing)
 SIMILAR < 0.5337 to the left, improve=2.644130, (0 missing)
 ZMUSICAC < -0.7153924 to the right, improve=2.572189, (0 missing)
 Surrogate splits:
 SIMILAR < 0.384 to the left, agree=0.864, adj=0.514, (0 split)
 NUMNOEND < 84.5 to the right, agree=0.770, adj=0.181, (0 split)
 LENTEMEL < 15.5 to the left, agree=0.770, adj=0.181, (0 split)
 MANPHMEL < 4.5 to the right, agree=0.770, adj=0.181, (0 split)
 NUMMEEND < 1.5 to the right, agree=0.763, adj=0.153, (0 split)
 MANSECEN < 6.234375 to the left, agree=0.728, adj=0.028, (0 split)

Node number 10: 57 observations, complexity param=0.008385744
 predicted class=2 expected loss=0.5438596
 class counts: 3 26 12 12 4 0
 probabilities: 0.053 0.456 0.211 0.211 0.070 0.000
 left son=20 (14 obs) right son=21 (43 obs)
 Primary splits:
 ZMUSICCO < 0.50904 to the right, improve=4.871714, (0 missing)
 MANINTER < 0.5 to the right, improve=2.474520, (0 missing)
 LENTEMEL < 25.5 to the left, improve=2.255183, (0 missing)
 DURTEMEL < 13.5 to the left, improve=1.859649, (0 missing)
 ZMUSICAC < -0.01528496 to the left, improve=1.617103, (0 missing)

Node number 11: 61 observations, complexity param=0.008385744
 predicted class=3 expected loss=0.4918033
 class counts: 3 14 31 9 3 1
 probabilities: 0.049 0.230 0.508 0.148 0.049 0.016
 left son=22 (14 obs) right son=23 (47 obs)
 Primary splits:
 SIMILAR < 0.4554 to the left, improve=3.186756, (0 missing)
 JOINT52 < 0.6886 to the left, improve=2.622276, (0 missing)
 LENTEMEL < 21 to the left, improve=2.587951, (0 missing)

NOTEREST < 0.5 to the left, improve=2.247546, (0 missing)
 NUMNOEND < 21 to the right, improve=1.817685, (0 missing)
 Surrogate splits:
 DURTEMEL < 9.5 to the left, agree=0.869, adj=0.429, (0 split)
 MANOTEPH < 12 to the right, agree=0.836, adj=0.286, (0 split)
 NOTEREST < 0.5 to the left, agree=0.803, adj=0.143, (0 split)

Node number 12: 13 observations
 predicted class=4 expected loss=0.07692308
 class counts: 0 0 1 12 0 0
 probabilities: 0.000 0.000 0.077 0.923 0.000 0.000

Node number 13: 168 observations, complexity param=0.00524109
 predicted class=4 expected loss=0.6130952
 class counts: 7 19 24 65 28 25
 probabilities: 0.042 0.113 0.143 0.387 0.167 0.149
 left son=26 (106 obs) right son=27 (62 obs)
 Primary splits:
 ZMUSICCO < -0.07205736 to the left, improve=3.900168, (0 missing)
 ZMUSICAC < 0.3710032 to the left, improve=2.944110, (0 missing)
 DURTEMEL < 9.5 to the left, improve=1.602591, (0 missing)
 SECSSEND < 34.85 to the right, improve=1.520326, (0 missing)
 LENTEMEL < 14 to the left, improve=1.517977, (0 missing)
 Surrogate splits:
 ZMUSICAC < -0.07323501 to the left, agree=0.875, adj=0.661, (0 split)

Node number 14: 18 observations
 predicted class=5 expected loss=0.5
 class counts: 1 0 6 2 9 0
 probabilities: 0.056 0.000 0.333 0.111 0.500 0.000

Node number 15: 60 observations, complexity param=0.01118099
 predicted class=6 expected loss=0.7333333
 class counts: 3 4 12 14 11 16
 probabilities: 0.050 0.067 0.200 0.233 0.183 0.267
 left son=30 (16 obs) right son=31 (44 obs)
 Primary splits:
 NUMNOEND < 33 to the right, improve=2.985606, (0 missing)
 NUMMEEND < 1 to the right, improve=2.985606, (0 missing)
 SECSSEND < 20 to the right, improve=2.985606, (0 missing)
 DURTEMEL < 15.5 to the left, improve=2.985606, (0 missing)
 LENTEMEL < 24 to the left, improve=2.985606, (0 missing)
 Surrogate splits:
 NUMMEEND < 1 to the right, agree=1, adj=1, (0 split)
 SECSSEND < 20 to the right, agree=1, adj=1, (0 split)
 DURTEMEL < 15.5 to the left, agree=1, adj=1, (0 split)
 LENTEMEL < 24 to the left, agree=1, adj=1, (0 split)
 POCOMEL < 1.5 to the left, agree=1, adj=1, (0 split)

Node number 18: 72 observations, complexity param=0.006289308
 predicted class=3 expected loss=0.625

class counts: 22 14 27 7 2 0
probabilities: 0.306 0.194 0.375 0.097 0.028 0.000
left son=36 (42 obs) right son=37 (30 obs)

Primary splits:

ZMUSICAC < -0.4922688 to the right, improve=2.2468250, (0 missing)
ZMUSICCO < -0.2750667 to the left, improve=1.4362250, (0 missing)
HARMFUNC < 0.5 to the right, improve=0.9252137, (0 missing)
CHANGERA < 0.5 to the right, improve=0.8179876, (0 missing)
DURTEMEL < 13.5 to the left, improve=0.8105735, (0 missing)

Surrogate splits:

ZMUSICCO < -0.6812446 to the right, agree=0.819, adj=0.567, (0 split)
DURTEMEL < 13.5 to the left, agree=0.611, adj=0.067, (0 split)
JOINT52 < 0.3134 to the right, agree=0.611, adj=0.067, (0 split)
MANPHMEL < 2.75 to the left, agree=0.597, adj=0.033, (0 split)
CHANGERA < 0.5 to the left, agree=0.597, adj=0.033, (0 split)

Node number 19: 185 observations, complexity param=0.006289308

predicted class=3 expected loss=0.6216216

class counts: 18 31 70 51 10 5
probabilities: 0.097 0.168 0.378 0.276 0.054 0.027
left son=38 (13 obs) right son=39 (172 obs)

Primary splits:

SIMILAR < 0.41425 to the left, improve=2.708998, (0 missing)
MANSECEN < 8.5875 to the right, improve=2.468796, (0 missing)
ZMUSICAC < -0.7153924 to the right, improve=2.427540, (0 missing)
ZMUSICCO < -0.6628636 to the left, improve=2.377138, (0 missing)
LENTEMEL < 31.5 to the left, improve=1.731388, (0 missing)

Node number 20: 14 observations

predicted class=2 expected loss=0.1428571

class counts: 2 12 0 0 0 0
probabilities: 0.143 0.857 0.000 0.000 0.000 0.000

Node number 21: 43 observations, complexity param=0.008385744

predicted class=2 expected loss=0.6744186

class counts: 1 14 12 12 4 0
probabilities: 0.023 0.326 0.279 0.279 0.093 0.000
left son=42 (25 obs) right son=43 (18 obs)

Primary splits:

LENTEMEL < 25.5 to the left, improve=3.811059, (0 missing)
MANINTER < 0.5 to the right, improve=3.444075, (0 missing)
DURTEMEL < 13.5 to the left, improve=2.954749, (0 missing)
JOINT52 < 0.61805 to the left, improve=2.098837, (0 missing)
MANPHMEL < 1.75 to the left, improve=1.996985, (0 missing)

Surrogate splits:

JOINT52 < 0.6886 to the left, agree=0.767, adj=0.444, (0 split)
MANINTER < 0.5 to the left, agree=0.744, adj=0.389, (0 split)
NOTEREST < 0.5 to the left, agree=0.721, adj=0.333, (0 split)
MANRHYTH < 0.5 to the right, agree=0.674, adj=0.222, (0 split)
SESEND < 20.05 to the right, agree=0.674, adj=0.222, (0 split)
DURTEMEL < 13.5 to the left, agree=0.674, adj=0.222, (0 split)

MANSECEN < 27	to the right, agree=0.674, adj=0.222, (0 split)
MANPHMEL < 1.75	to the left, agree=0.674, adj=0.222, (0 split)
SIMILAR < 0.7235	to the right, agree=0.651, adj=0.167, (0 split)
MANOTEPH < 8.75	to the left, agree=0.628, adj=0.111, (0 split)
NUMNOEND < 11	to the right, agree=0.605, adj=0.056, (0 split)
NUMMEEND < 0.55	to the right, agree=0.605, adj=0.056, (0 split)
POCOMEL < 1.5	to the left, agree=0.605, adj=0.056, (0 split)

Node number 22: 14 observations

predicted class=2 expected loss=0.4285714

class counts: 0 8 4 1 1 0

probabilities: 0.000 0.571 0.286 0.071 0.071 0.000

Node number 23: 47 observations

predicted class=3 expected loss=0.4255319

class counts: 3 6 27 8 2 1

probabilities: 0.064 0.128 0.574 0.170 0.043 0.021

Node number 26: 106 observations

predicted class=4 expected loss=0.5377358

class counts: 6 12 17 49 9 13

probabilities: 0.057 0.113 0.160 0.462 0.085 0.123

Node number 27: 62 observations, complexity param=0.00524109

predicted class=5 expected loss=0.6935484

class counts: 1 7 7 16 19 12

probabilities: 0.016 0.113 0.113 0.258 0.306 0.194

left son=54 (35 obs) right son=55 (27 obs)

Primary splits:

ZMUSICCO < 0.338823 to the right, improve=2.031678, (0 missing)

ZMUSICAC < 0.3121117 to the left, improve=2.012630, (0 missing)

SESEND < 36.75 to the right, improve=1.648655, (0 missing)

LENTEMEL < 14 to the left, improve=1.648655, (0 missing)

DURTEMEL < 9.5 to the left, improve=1.644417, (0 missing)

Surrogate splits:

ZMUSICAC < 0.137027 to the left, agree=0.758, adj=0.444, (0 split)

DURTEMEL < 11 to the right, agree=0.613, adj=0.111, (0 split)

LENTEMEL < 18 to the right, agree=0.597, adj=0.074, (0 split)

SIMILAR < 0.8779 to the right, agree=0.581, adj=0.037, (0 split)

MANRHYTH < 0.5 to the left, agree=0.581, adj=0.037, (0 split)

JOINT52 < 0.9706 to the right, agree=0.581, adj=0.037, (0 split)

Node number 30: 16 observations

predicted class=3 expected loss=0.5625

class counts: 2 1 7 4 2 0

probabilities: 0.125 0.063 0.438 0.250 0.125 0.000

Node number 31: 44 observations

predicted class=6 expected loss=0.6363636

class counts: 1 3 5 10 9 16

probabilities: 0.023 0.068 0.114 0.227 0.205 0.364

Node number 36: 42 observations, complexity param=0.006289308

predicted class=1 expected loss=0.5714286

class counts: 18 5 16 3 0 0

probabilities: 0.429 0.119 0.381 0.071 0.000 0.000

left son=72 (22 obs) right son=73 (20 obs)

Primary splits:

ZMUSICCO < -0.2642584 to the left, improve=1.8264070, (0 missing)

ZMUSICAC < -0.005839524 to the left, improve=1.2264070, (0 missing)

HARMFUNC < 0.5 to the right, improve=0.8875837, (0 missing)

SECSSEND < 17 to the right, improve=0.8082251, (0 missing)

MANSECEN < 21.90625 to the right, improve=0.8082251, (0 missing)

Surrogate splits:

ZMUSICAC < -0.005839524 to the left, agree=0.810, adj=0.60, (0 split)

CHANGERA < 0.5 to the right, agree=0.619, adj=0.20, (0 split)

LENTEMEL < 17.5 to the right, agree=0.595, adj=0.15, (0 split)

HARMFUNC < 0.5 to the right, agree=0.595, adj=0.15, (0 split)

SIMILAR < 0.3366 to the right, agree=0.571, adj=0.10, (0 split)

MANRHYTH < 0.5 to the left, agree=0.571, adj=0.10, (0 split)

MANCONT < 0.5 to the right, agree=0.571, adj=0.10, (0 split)

SECSSEND < 17 to the left, agree=0.571, adj=0.10, (0 split)

JOINT52 < 0.4818 to the right, agree=0.571, adj=0.10, (0 split)

MANSECEN < 21.90625 to the left, agree=0.571, adj=0.10, (0 split)

MANOTEPH < 6 to the right, agree=0.571, adj=0.10, (0 split)

MANPHMEL < 1.25 to the right, agree=0.571, adj=0.10, (0 split)

NUMNOEND < 9.5 to the left, agree=0.548, adj=0.05, (0 split)

NUMMEEND < 0.5 to the left, agree=0.548, adj=0.05, (0 split)

DURTEMEL < 10.5 to the right, agree=0.548, adj=0.05, (0 split)

POCOMEL < 1.5 to the right, agree=0.548, adj=0.05, (0 split)

Node number 37: 30 observations

predicted class=3 expected loss=0.6333333

class counts: 4 9 11 4 2 0

probabilities: 0.133 0.300 0.367 0.133 0.067 0.000

Node number 38: 13 observations

predicted class=4 expected loss=0.6153846

class counts: 3 4 0 5 1 0

probabilities: 0.231 0.308 0.000 0.385 0.077 0.000

Node number 39: 172 observations

predicted class=3 expected loss=0.5930233

class counts: 15 27 70 46 9 5

probabilities: 0.087 0.157 0.407 0.267 0.052 0.029

Node number 42: 25 observations

predicted class=2 expected loss=0.48

class counts: 0 13 3 7 2 0

probabilities: 0.000 0.520 0.120 0.280 0.080 0.000

Node number 43: 18 observations

predicted class=3 expected loss=0.5
class counts: 1 1 9 5 2 0
probabilities: 0.056 0.056 0.500 0.278 0.111 0.000

Node number 54: 35 observations
predicted class=5 expected loss=0.6571429
class counts: 0 6 5 10 12 2
probabilities: 0.000 0.171 0.143 0.286 0.343 0.057

Node number 55: 27 observations, complexity param=0.00524109
predicted class=6 expected loss=0.6296296
class counts: 1 1 2 6 7 10
probabilities: 0.037 0.037 0.074 0.222 0.259 0.370
left son=110 (7 obs) right son=111 (20 obs)

Primary splits:
ZMUSICAC < 1.474184 to the right, improve=2.525926, (0 missing)
ZMUSICCO < 0.1411808 to the left, improve=1.849003, (0 missing)
NUMNOEND < 14.5 to the right, improve=1.739113, (0 missing)
NUMMEEND < 0.5 to the right, improve=1.739113, (0 missing)
SECSSEND < 16 to the right, improve=1.739113, (0 missing)

Node number 72: 22 observations
predicted class=1 expected loss=0.4090909
class counts: 13 2 6 1 0 0
probabilities: 0.591 0.091 0.273 0.045 0.000 0.000

Node number 73: 20 observations
predicted class=3 expected loss=0.5
class counts: 5 3 10 2 0 0
probabilities: 0.250 0.150 0.500 0.100 0.000 0.000

Node number 110: 7 observations
predicted class=4 expected loss=0.4285714
class counts: 0 0 1 4 2 0
probabilities: 0.000 0.000 0.143 0.571 0.286 0.000

Node number 111: 20 observations
predicted class=6 expected loss=0.5
class counts: 1 1 1 2 5 10
probabilities: 0.050 0.050 0.050 0.100 0.250 0.500

A-5.6.3.3 Regressionsbaum

Baumzusammenfassung:

- 1) root 665 1235.4200 3.306767
- 2) SIMILAR < 0.86035 406 567.2241 2.844828
- 4) JOINT52 < 0.61805 109 140.9725 2.449541 *
- 5) JOINT52 >= 0.61805 297 402.9697 2.989899

- 10) SIMILAR < 0.4926 93 121.2903 2.645161 *
- 11) SIMILAR >= 0.4926 204 265.5882 3.147059 *
- 3) SIMILAR >= 0.86035 259 445.7529 4.030888
- 6) LENTEMEL < 16.5 73 118.2466 3.493151
- 12) ZMUSICAC < 0.3121117 57 88.5614 3.245614 *
- 13) ZMUSICAC >= 0.3121117 16 13.7500 4.375000 *
- 7) LENTEMEL >= 16.5 186 298.1129 4.241935 *

Regressionsbaum: Vollständige Beschreibung:

```
rpart(formula = RATDACOM ~ SIMILAR + MANRHYTH + MANINTER +
MANCONT +
NUMNOEND + NUMMEEND + SECSSEND + DURTEMEL + LENTEMEL +
POCOMEL +
JOINT52 + MANSECEN + MANOTEPH + MANPHMEL + NOTEREST +
HARMFUNC +
CHANGERA + ZMUSICCO + ZMUSICAC, data = treedata, method = "anova",
control = list(maxsurrogate = 18, cp = 0.005))
n= 665
```

	CP	nsplit	rel error	xerror	xstd
1	0.180054230	0	1.0000000	1.0039323	0.04666905
2	0.023792255	1	0.8199458	0.8524988	0.04261630
3	0.018845391	2	0.7961535	0.8599105	0.04322279
4	0.013024838	3	0.7773081	0.8323163	0.04280122
5	0.012898591	4	0.7642833	0.8291916	0.04417916
6	0.008928146	5	0.7513847	0.8223412	0.04380699
7	0.008106478	8	0.7246003	0.8330975	0.04469608
8	0.007537708	9	0.7164938	0.8275506	0.04412771
9	0.006715350	11	0.7014184	0.8308071	0.04413912
10	0.006294479	13	0.6879877	0.8261237	0.04439083
11	0.005977844	18	0.6565153	0.8319797	0.04523014
12	0.005821396	20	0.6445596	0.8326739	0.04573118
13	0.005254398	22	0.6329168	0.8369670	0.04553796
14	0.005000000	24	0.6224080	0.8376995	0.04562572

Node number 1: 665 observations, complexity param=0.1800542

mean=3.306767, MSE=1.857774

left son=2 (406 obs) right son=3 (259 obs)

Primary splits:

SIMILAR < 0.86035 to the left, improve=0.18005420, (0 missing)

JOINT52 < 0.91795 to the left, improve=0.18005420, (0 missing)

HARMFUNC < 0.5 to the right, improve=0.07637188, (0 missing)

LENTEMEL < 32.5 to the left, improve=0.07082162, (0 missing)

CHANGERA < 0.5 to the right, improve=0.07029454, (0 missing)

Surrogate splits:

JOINT52 < 0.91795 to the left, agree=1.000, adj=1.000, (0 split)

LENTEMEL < 32.5 to the left, agree=0.726, adj=0.297, (0 split)

CHANGERA < 0.5 to the right, agree=0.689, adj=0.201, (0 split)

SECSSEND < 30 to the left, agree=0.674, adj=0.162, (0 split)

HARMFUNC < 0.5 to the right, agree=0.674, adj=0.162, (0 split)
 NOTEREST < 0.5 to the right, agree=0.662, adj=0.131, (0 split)
 DURTEMEL < 15.5 to the left, agree=0.645, adj=0.089, (0 split)
 MANCONT < 0.5 to the right, agree=0.618, adj=0.019, (0 split)

Node number 2: 406 observations, complexity param=0.01884539

mean=2.844828, MSE=1.397104

left son=4 (109 obs) right son=5 (297 obs)

Primary splits:

JOINT52 < 0.61805 to the left, improve=0.04104544, (0 missing)
 SIMILAR < 0.50105 to the left, improve=0.03611823, (0 missing)
 LENTEMEL < 27.5 to the left, improve=0.02291764, (0 missing)
 ZMUSICAC < -0.4523218 to the right, improve=0.02218653, (0 missing)
 NUMNOEND < 84.5 to the right, improve=0.01790474, (0 missing)

Surrogate splits:

SIMILAR < 0.384 to the left, agree=0.860, adj=0.477, (0 split)
 NUMNOEND < 84.5 to the right, agree=0.781, adj=0.183, (0 split)
 NUMMEEND < 1.5 to the right, agree=0.776, adj=0.165, (0 split)
 MANPHMEL < 3.25 to the right, agree=0.776, adj=0.165, (0 split)
 LENTEMEL < 15.5 to the left, agree=0.773, adj=0.156, (0 split)
 DURTEMEL < 8.5 to the left, agree=0.739, adj=0.028, (0 split)
 SECSSEND < 31.8 to the right, agree=0.734, adj=0.009, (0 split)
 MANSECEN < 35.625 to the right, agree=0.734, adj=0.009, (0 split)

Node number 3: 259 observations, complexity param=0.02379226

mean=4.030888, MSE=1.721054

left son=6 (73 obs) right son=7 (186 obs)

Primary splits:

LENTEMEL < 16.5 to the left, improve=0.06594106, (0 missing)
 SECSSEND < 36.75 to the right, improve=0.05814967, (0 missing)
 DURTEMEL < 11 to the left, improve=0.03044291, (0 missing)
 ZMUSICAC < 0.2085403 to the left, improve=0.03028930, (0 missing)
 NUMNOEND < 40 to the right, improve=0.02915991, (0 missing)

Surrogate splits:

SECSSEND < 36.75 to the right, agree=0.934, adj=0.767, (0 split)
 NUMNOEND < 58.5 to the right, agree=0.795, adj=0.274, (0 split)
 NUMMEEND < 1.5 to the right, agree=0.795, adj=0.274, (0 split)
 DURTEMEL < 11 to the left, agree=0.784, adj=0.233, (0 split)
 POCOMEL < 1.5 to the left, agree=0.722, adj=0.014, (0 split)

Node number 4: 109 observations, complexity param=0.005821396

mean=2.449541, MSE=1.293325

left son=8 (43 obs) right son=9 (66 obs)

Primary splits:

ZMUSICCO < -0.2750667 to the left, improve=0.04142145, (0 missing)
 ZMUSICAC < -0.4523218 to the right, improve=0.02862140, (0 missing)
 HARMFUNC < 0.5 to the right, improve=0.02081012, (0 missing)
 JOINT52 < 0.4818 to the right, improve=0.01805057, (0 missing)
 SIMILAR < 0.5161 to the right, improve=0.01558998, (0 missing)

Surrogate splits:

ZMUSICAC < -0.07323501 to the left, agree=0.78, adj=0.442, (0 split)

Node number 5: 297 observations, complexity param=0.01302484

mean=2.989899, MSE=1.3568

left son=10 (93 obs) right son=11 (204 obs)

Primary splits:

SIMILAR < 0.4926 to the left, improve=0.03993139, (0 missing)

HARMFUNC < 0.5 to the right, improve=0.02339920, (0 missing)

ZMUSICAC < -0.5375717 to the right, improve=0.02124152, (0 missing)

ZMUSICCO < 0.1411808 to the right, improve=0.01549118, (0 missing)

MANOTEPH < 4.5 to the right, improve=0.01546287, (0 missing)

Surrogate splits:

MANOTEPH < 12 to the right, agree=0.751, adj=0.204, (0 split)

HARMFUNC < 0.5 to the right, agree=0.751, adj=0.204, (0 split)

CHANGERA < 0.5 to the right, agree=0.707, adj=0.065, (0 split)

MANSECEN < 8.2325 to the left, agree=0.697, adj=0.032, (0 split)

DURTEMEL < 9.5 to the left, agree=0.694, adj=0.022, (0 split)

JOINT52 < 0.66875 to the left, agree=0.694, adj=0.022, (0 split)

Node number 6: 73 observations, complexity param=0.01289859

mean=3.493151, MSE=1.619816

left son=12 (57 obs) right son=13 (16 obs)

Primary splits:

ZMUSICAC < 0.3121117 to the left, improve=0.13476220, (0 missing)

DURTEMEL < 9.5 to the left, improve=0.06991817, (0 missing)

NUMNOEND < 35 to the right, improve=0.06951194, (0 missing)

ZMUSICCO < 0.1411808 to the left, improve=0.04607323, (0 missing)

SECSSEND < 37.75 to the left, improve=0.02132058, (0 missing)

Surrogate splits:

ZMUSICCO < -1.250216 to the right, agree=0.808, adj=0.125, (0 split)

Node number 7: 186 observations, complexity param=0.008928146

mean=4.241935, MSE=1.602758

left son=14 (17 obs) right son=15 (169 obs)

Primary splits:

JOINT52 < 0.95895 to the left, improve=0.02681956, (0 missing)

ZMUSICAC < -1.861554 to the right, improve=0.02658240, (0 missing)

MANSECEN < 7.55 to the right, improve=0.02359643, (150 missing)

MANOTEPH < 2 to the right, improve=0.02359643, (150 missing)

MANPHMEL < 3.25 to the left, improve=0.02359643, (150 missing)

Node number 8: 43 observations, complexity param=0.005821396

mean=2.162791, MSE=1.252569

left son=16 (22 obs) right son=17 (21 obs)

Primary splits:

ZMUSICCO < -0.82861 to the right, improve=0.15864040, (0 missing)

ZMUSICAC < -0.5167185 to the right, improve=0.12725420, (0 missing)

JOINT52 < 0.4818 to the right, improve=0.04882423, (0 missing)

HARMFUNC < 0.5 to the right, improve=0.03468513, (0 missing)

SIMILAR < 0.5161 to the right, improve=0.03122687, (0 missing)

Surrogate splits:

ZMUSICAC < -0.5167185 to the right, agree=0.814, adj=0.619, (0 split)

CHANGERA < 0.5 to the left, agree=0.605, adj=0.190, (0 split)
 DURTEMEL < 12 to the left, agree=0.581, adj=0.143, (0 split)
 JOINT52 < 0.3134 to the right, agree=0.581, adj=0.143, (0 split)
 MANPHMEL < 2.75 to the left, agree=0.581, adj=0.143, (0 split)
 SIMILAR < 0.506 to the left, agree=0.558, adj=0.095, (0 split)
 LENTEMEL < 22 to the right, agree=0.558, adj=0.095, (0 split)
 MANOTEPH < 3.5 to the right, agree=0.558, adj=0.095, (0 split)
 NUMNOEND < 9.5 to the right, agree=0.535, adj=0.048, (0 split)
 NUMMEEND < 0.5 to the right, agree=0.535, adj=0.048, (0 split)
 SECSSEND < 8 to the right, agree=0.535, adj=0.048, (0 split)
 POCOMEL < 1.5 to the left, agree=0.535, adj=0.048, (0 split)
 MANSECEN < 7.509375 to the left, agree=0.535, adj=0.048, (0 split)

Node number 9: 66 observations
 mean=2.636364, MSE=1.231405

Node number 10: 93 observations, complexity param=0.007537708
 mean=2.645161, MSE=1.304197
 left son=20 (59 obs) right son=21 (34 obs)

Primary splits:

ZMUSICAC < -0.4523218 to the right, improve=0.07560934, (0 missing)
 ZMUSICCO < -0.1463046 to the right, improve=0.04546654, (0 missing)
 NUMNOEND < 11.5 to the right, improve=0.04057162, (0 missing)
 NUMMEEND < 0.5 to the right, improve=0.04057162, (0 missing)
 SECSSEND < 8.5 to the right, improve=0.04057162, (0 missing)

Surrogate splits:

ZMUSICCO < -0.6812446 to the right, agree=0.860, adj=0.618, (0 split)
 HARMFUNC < 0.5 to the right, agree=0.656, adj=0.059, (0 split)

Node number 11: 204 observations, complexity param=0.008106478
 mean=3.147059, MSE=1.301903
 left son=22 (54 obs) right son=23 (150 obs)

Primary splits:

DURTEMEL < 10.5 to the left, improve=0.03770838, (0 missing)
 MANOTEPH < 6.25 to the right, improve=0.02595404, (0 missing)
 JOINT52 < 0.81175 to the right, improve=0.02500112, (0 missing)
 MANSECEN < 6.375 to the left, improve=0.02135137, (0 missing)
 SIMILAR < 0.5301 to the right, improve=0.01795507, (0 missing)

Surrogate splits:

JOINT52 < 0.81175 to the right, agree=0.912, adj=0.667, (0 split)
 NUMNOEND < 53 to the right, agree=0.828, adj=0.352, (0 split)
 MANSECEN < 6.375 to the left, agree=0.824, adj=0.333, (0 split)
 SECSSEND < 38.125 to the right, agree=0.819, adj=0.315, (0 split)
 SIMILAR < 0.7235 to the right, agree=0.814, adj=0.296, (0 split)
 MANRHYTH < 0.5 to the right, agree=0.814, adj=0.296, (0 split)
 LENTEMEL < 27 to the left, agree=0.814, adj=0.296, (0 split)

Node number 12: 57 observations
 mean=3.245614, MSE=1.553709

Node number 13: 16 observations

mean=4.375, MSE=0.859375

Node number 14: 17 observations
mean=3.588235, MSE=1.301038

Node number 15: 169 observations, complexity param=0.008928146
mean=4.307692, MSE=1.585799
left son=30 (162 obs) right son=31 (7 obs)

Primary splits:

ZMUSICAC < -1.861554 to the right, improve=0.02606347, (0 missing)
DURTEMEL < 13.5 to the left, improve=0.01529586, (0 missing)
ZMUSICCO < 0.484234 to the right, improve=0.01298450, (0 missing)
LENTEMEL < 21.5 to the left, improve=0.01182071, (0 missing)
NOTEREST < 0.5 to the right, improve=0.01017692, (0 missing)

Node number 16: 22 observations
mean=1.727273, MSE=0.9256198

Node number 17: 21 observations
mean=2.619048, MSE=1.188209

Node number 20: 59 observations, complexity param=0.005254398
mean=2.40678, MSE=1.020971
left son=40 (33 obs) right son=41 (26 obs)

Primary splits:

ZMUSICCO < 0.4230768 to the left, improve=0.081004300, (0 missing)
ZMUSICAC < 0.3710032 to the right, improve=0.058106660, (0 missing)
MANSECEN < 8.9575 to the right, improve=0.012219970, (0 missing)
DURTEMEL < 11 to the right, improve=0.007795132, (0 missing)
MANOTEPH < 9.125 to the left, improve=0.007795132, (0 missing)

Surrogate splits:

ZMUSICAC < 1.132064 to the left, agree=0.678, adj=0.269, (0 split)

Node number 21: 34 observations, complexity param=0.007537708
mean=3.058824, MSE=1.525952
left son=42 (14 obs) right son=43 (20 obs)

Primary splits:

NUMNOEND < 11.5 to the right, improve=0.1822157, (0 missing)
NUMMEEND < 0.5 to the right, improve=0.1822157, (0 missing)
SECSSEND < 8.5 to the right, improve=0.1822157, (0 missing)
POCOMEL < 1.5 to the left, improve=0.1822157, (0 missing)
MANSECEN < 17.5625 to the right, improve=0.1822157, (0 missing)

Surrogate splits:

NUMMEEND < 0.5 to the right, agree=1.000, adj=1.000, (0 split)
SECSSEND < 8.5 to the right, agree=1.000, adj=1.000, (0 split)
POCOMEL < 1.5 to the left, agree=1.000, adj=1.000, (0 split)
MANSECEN < 17.5625 to the right, agree=1.000, adj=1.000, (0 split)
SIMILAR < 0.4429 to the right, agree=0.794, adj=0.500, (0 split)
DURTEMEL < 14.5 to the right, agree=0.794, adj=0.500, (0 split)
LENTEMEL < 21 to the left, agree=0.794, adj=0.500, (0 split)
JOINT52 < 0.68705 to the left, agree=0.794, adj=0.500, (0 split)

MANOTEPH < 4 to the right, agree=0.794, adj=0.500, (0 split)
 MANPHMEL < 2.25 to the left, agree=0.794, adj=0.500, (0 split)
 NOTEREST < 0.5 to the left, agree=0.794, adj=0.500, (0 split)
 CHANGERA < 0.5 to the left, agree=0.794, adj=0.500, (0 split)
 HARMFUNC < 0.5 to the left, agree=0.706, adj=0.286, (0 split)

Node number 22: 54 observations
 mean=2.777778, MSE=1.061728

Node number 23: 150 observations, complexity param=0.00671535
 mean=3.28, MSE=1.3216

left son=46 (121 obs) right son=47 (29 obs)

Primary splits:

ZMUSICAC < -0.7153924 to the right, improve=0.041542780, (0 missing)
 HARMFUNC < 0.5 to the right, improve=0.016395570, (0 missing)
 MANOTEPH < 7.25 to the right, improve=0.013468980, (0 missing)
 ZMUSICCO < -1.042813 to the right, improve=0.012569870, (0 missing)
 SIMILAR < 0.5301 to the right, improve=0.008761509, (0 missing)

Surrogate splits:

ZMUSICCO < -0.6812446 to the right, agree=0.893, adj=0.448, (0 split)

Node number 30: 162 observations, complexity param=0.008928146
 mean=4.265432, MSE=1.602385

left son=60 (12 obs) right son=61 (150 obs)

Primary splits:

ZMUSICAC < -1.156763 to the left, improve=0.069763870, (0 missing)
 ZMUSICCO < -0.4802553 to the left, improve=0.016910890, (0 missing)
 NOTEREST < 0.5 to the right, improve=0.014713600, (0 missing)
 DURTEMEL < 13.5 to the left, improve=0.014126220, (0 missing)
 LENTEMEL < 21.5 to the left, improve=0.009363208, (0 missing)

Node number 31: 7 observations
 mean=5.285714, MSE=0.2040816

Node number 40: 33 observations, complexity param=0.005254398
 mean=2.151515, MSE=0.9770432

left son=80 (10 obs) right son=81 (23 obs)

Primary splits:

ZMUSICCO < 0.1411808 to the right, improve=0.25132400, (0 missing)
 ZMUSICAC < 0.2085403 to the right, improve=0.14915410, (0 missing)
 LENTEMEL < 27.5 to the right, improve=0.04060150, (0 missing)
 MANINTER < 0.5 to the right, improve=0.01933405, (0 missing)
 MANCONT < 0.5 to the left, improve=0.01933405, (0 missing)

Surrogate splits:

ZMUSICAC < 0.3166205 to the right, agree=0.818, adj=0.4, (0 split)

Node number 41: 26 observations
 mean=2.730769, MSE=0.8890533

Node number 42: 14 observations
 mean=2.428571, MSE=0.9591837

Node number 43: 20 observations
mean=3.5, MSE=1.45

Node number 46: 121 observations, complexity param=0.00671535
mean=3.165289, MSE=1.212349
left son=92 (15 obs) right son=93 (106 obs)

Primary splits:

ZMUSICCO < -0.6628636 to the left, improve=0.05696958, (0 missing)
MANOTEPH < 7.25 to the right, improve=0.03093763, (0 missing)
ZMUSICAC < 0.05211053 to the left, improve=0.02940495, (0 missing)
DURTEMEL < 12.5 to the left, improve=0.02569215, (0 missing)
SIMILAR < 0.5301 to the right, improve=0.02074178, (0 missing)

Surrogate splits:

ZMUSICAC < -0.6019683 to the left, agree=0.901, adj=0.2, (0 split)

Node number 47: 29 observations
mean=3.758621, MSE=1.49346

Node number 60: 12 observations
mean=3.083333, MSE=1.576389

Node number 61: 150 observations, complexity param=0.006294479
mean=4.36, MSE=1.483733
left son=122 (135 obs) right son=123 (15 obs)

Primary splits:

ZMUSICAC < -0.7153924 to the right, improve=0.024615920, (0 missing)
ZMUSICCO < 0.484234 to the right, improve=0.023552380, (0 missing)
NOTEREST < 0.5 to the right, improve=0.014169960, (0 missing)
DURTEMEL < 13.5 to the left, improve=0.014074900, (0 missing)
LENTEMEL < 21.5 to the left, improve=0.008607369, (0 missing)

Node number 80: 10 observations
mean=1.4, MSE=0.64

Node number 81: 23 observations
mean=2.478261, MSE=0.7712665

Node number 92: 15 observations
mean=2.466667, MSE=1.448889

Node number 93: 106 observations
mean=3.264151, MSE=1.100036

Node number 122: 135 observations, complexity param=0.006294479
mean=4.296296, MSE=1.527023
left son=244 (13 obs) right son=245 (122 obs)

Primary splits:

ZMUSICAC < -0.5375717 to the left, improve=0.040076230, (0 missing)
ZMUSICCO < -0.07205736 to the left, improve=0.016904420, (0 missing)
NOTEREST < 0.5 to the right, improve=0.014305170, (0 missing)

DURTEMEL < 13.5 to the left, improve=0.013129290, (0 missing)
LENTEMEL < 21.5 to the left, improve=0.008533347, (0 missing)

Node number 123: 15 observations
mean=4.933333, MSE=0.7288889

Node number 244: 13 observations
mean=3.538462, MSE=0.556213

Node number 245: 122 observations, complexity param=0.006294479
mean=4.377049, MSE=1.562752
left son=490 (40 obs) right son=491 (82 obs)

Primary splits:

ZMUSICCO < 0.484234 to the right, improve=0.033387370, (0 missing)
NOTEREST < 0.5 to the right, improve=0.013559290, (0 missing)
ZMUSICAC < 1.132064 to the right, improve=0.008719279, (0 missing)
DURTEMEL < 13.5 to the left, improve=0.008533840, (0 missing)
LENTEMEL < 19.5 to the left, improve=0.005690548, (0 missing)

Surrogate splits:

ZMUSICAC < 1.132064 to the right, agree=0.746, adj=0.225, (0 split)

Node number 490: 40 observations
mean=4.05, MSE=0.9475

Node number 491: 82 observations, complexity param=0.006294479
mean=4.536585, MSE=1.785247
left son=982 (75 obs) right son=983 (7 obs)

Primary splits:

ZMUSICCO < 0.4230768 to the left, improve=0.05598705, (0 missing)
ZMUSICAC < 0.2085403 to the left, improve=0.04268000, (0 missing)
NOTEREST < 0.5 to the right, improve=0.02358385, (0 missing)
LENTEMEL < 19.5 to the left, improve=0.02354429, (0 missing)
DURTEMEL < 13.5 to the left, improve=0.02061339, (0 missing)

Node number 982: 75 observations, complexity param=0.006294479
mean=4.44, MSE=1.819733
left son=1964 (45 obs) right son=1965 (30 obs)

Primary splits:

ZMUSICAC < 0.2085403 to the left, improve=0.07752052, (0 missing)
NOTEREST < 0.5 to the right, improve=0.03164794, (0 missing)
ZMUSICCO < -0.07205736 to the left, improve=0.02462968, (0 missing)
NUMNOEND < 19.5 to the left, improve=0.02214484, (0 missing)
NUMMEEND < 0.5 to the left, improve=0.02214484, (0 missing)

Surrogate splits:

ZMUSICCO < -0.07205736 to the left, agree=0.84, adj=0.6, (0 split)

Node number 983: 7 observations
mean=5.571429, MSE=0.244898

Node number 1964: 45 observations, complexity param=0.005977844
mean=4.133333, MSE=1.671111

left son=3928 (23 obs) right son=3929 (22 obs)

Primary splits:

ZMUSICCO < -0.2642584 to the right, improve=0.07695421, (0 missing)

ZMUSICAC < -0.1317757 to the right, improve=0.07111899, (0 missing)

DURTEMEL < 13.5 to the left, improve=0.03708868, (0 missing)

LENTEMEL < 28 to the left, improve=0.03404255, (0 missing)

NOTEREST < 0.5 to the right, improve=0.02606383, (0 missing)

Surrogate splits:

ZMUSICAC < -0.04162499 to the right, agree=0.867, adj=0.727, (0 split)

SIMILAR < 0.8768 to the right, agree=0.533, adj=0.045, (0 split)

MANRHYTH < 0.5 to the left, agree=0.533, adj=0.045, (0 split)

NUMNOEND < 19.5 to the right, agree=0.533, adj=0.045, (0 split)

NUMMEEND < 0.5 to the right, agree=0.533, adj=0.045, (0 split)

SECSSEND < 16 to the right, agree=0.533, adj=0.045, (0 split)

DURTEMEL < 15 to the left, agree=0.533, adj=0.045, (0 split)

LENTEMEL < 19.5 to the right, agree=0.533, adj=0.045, (0 split)

POCOMEL < 1.5 to the left, agree=0.533, adj=0.045, (0 split)

JOINT52 < 0.98835 to the right, agree=0.533, adj=0.045, (0 split)

Node number 1965: 30 observations

mean=4.9, MSE=1.69

Node number 3928: 23 observations

mean=3.782609, MSE=0.6049149

Node number 3929: 22 observations, complexity param=0.005977844

mean=4.5, MSE=2.522727

left son=7858 (10 obs) right son=7859 (12 obs)

Primary splits:

LENTEMEL < 19.5 to the left, improve=0.16186190, (0 missing)

ZMUSICCO < -0.4802553 to the left, improve=0.16186190, (0 missing)

ZMUSICAC < -0.1317757 to the right, improve=0.16186190, (0 missing)

DURTEMEL < 15 to the left, improve=0.06860707, (0 missing)

NOTEREST < 0.5 to the right, improve=0.01321321, (0 missing)

Surrogate splits:

DURTEMEL < 15 to the left, agree=0.864, adj=0.7, (0 split)

ZMUSICCO < -0.3108498 to the left, agree=0.682, adj=0.3, (0 split)

SIMILAR < 0.8768 to the right, agree=0.636, adj=0.2, (0 split)

MANRHYTH < 0.5 to the left, agree=0.636, adj=0.2, (0 split)

SECSSEND < 32.85 to the right, agree=0.636, adj=0.2, (0 split)

JOINT52 < 0.98835 to the right, agree=0.636, adj=0.2, (0 split)

ZMUSICAC < -0.1317757 to the right, agree=0.636, adj=0.2, (0 split)

NUMNOEND < 19.5 to the right, agree=0.591, adj=0.1, (0 split)

NUMMEEND < 0.5 to the right, agree=0.591, adj=0.1, (0 split)

POCOMEL < 1.5 to the left, agree=0.591, adj=0.1, (0 split)

Node number 7858: 10 observations

mean=3.8, MSE=2.96

Node number 7859: 12 observations

mean=5.083333, MSE=1.409722

A-5.6.3.4 2-stufiger Klassifizierungsbaum:

1. Stufe:

Baumzusammenfassung:

```
> tree1class_2step  
n= 665
```

```
node), split, n, loss, yval, (yprob)  
* denotes terminal node
```

- 1) root 665 294 0 (0.5578947 0.4421053)
- 2) SIMILAR < 0.86035 406 112 0 (0.7241379 0.2758621) *
- 3) SIMILAR >= 0.86035 259 77 1 (0.2972973 0.7027027)
- 6) SECSSEND >= 36.75 56 27 0 (0.5178571 0.4821429)
- 12) NUMNOEND >= 35 37 13 0 (0.6486486 0.3513514) *
- 13) NUMNOEND < 35 19 5 1 (0.2631579 0.7368421) *
- 7) SECSSEND < 36.75 203 48 1 (0.2364532 0.7635468) *

1. Stufe zweistufiger Klassifizierungsbaum: Vollständige Beschreibung:

```
rpart(formula = IDENT ~ SIMILAR + MANRHYTH + MANINTER + MANCONT +  
  NUMNOEND + NUMMEEND + SECSSEND + DURTEMEL + LENTEMEL +  
  POCOMEL +  
  JOINT52 + MANSECEN + MANOTEPH + MANPHMEL + NOTEREST +  
  HARMFUNC +  
  CHANGERA + ZMUSICCO + ZMUSICAC, data = treedata_2step, method = "class",  
  control = list(maxsurrogate = 18, cp = 0.01))  
n= 665
```

	CP	nsplit	rel error	xerror	xstd
1	0.35714286	0	1.0000000	1.0000000	0.04356146
2	0.01870748	1	0.6428571	0.6530612	0.03974871
3	0.01000000	3	0.6054422	0.6700680	0.04004957

```
Node number 1: 665 observations, complexity param=0.3571429  
predicted class=0 expected loss=0.4421053
```

```
class counts: 371 294
```

```
probabilities: 0.558 0.442
```

```
left son=2 (406 obs) right son=3 (259 obs)
```

```
Primary splits:
```

```
SIMILAR < 0.86035 to the left, improve=57.61899, (0 missing)
```

```
JOINT52 < 0.91795 to the left, improve=57.61899, (0 missing)
```

```
CHANGERA < 0.5 to the right, improve=21.79078, (0 missing)
```

```
HARMFUNC < 0.5 to the right, improve=20.85381, (0 missing)
```

```
LENTEMEL < 32.5 to the left, improve=15.68096, (0 missing)
```

```
Surrogate splits:
```

```
JOINT52 < 0.91795 to the left, agree=1.000, adj=1.000, (0 split)
```

LENTEMEL < 32.5 to the left, agree=0.726, adj=0.297, (0 split)
 CHANGERA < 0.5 to the right, agree=0.689, adj=0.201, (0 split)
 SECSSEND < 30 to the left, agree=0.674, adj=0.162, (0 split)
 HARMFUNC < 0.5 to the right, agree=0.674, adj=0.162, (0 split)
 NOTEREST < 0.5 to the right, agree=0.662, adj=0.131, (0 split)
 DURTEMEL < 15.5 to the left, agree=0.645, adj=0.089, (0 split)
 MANCONT < 0.5 to the right, agree=0.618, adj=0.019, (0 split)

Node number 2: 406 observations

predicted class=0 expected loss=0.2758621

class counts: 294 112

probabilities: 0.724 0.276

Node number 3: 259 observations, complexity param=0.01870748

predicted class=1 expected loss=0.2972973

class counts: 77 182

probabilities: 0.297 0.703

left son=6 (56 obs) right son=7 (203 obs)

Primary splits:

SECSSEND < 36.75 to the right, improve=6.951438, (0 missing)
 LENTEMEL < 15.5 to the left, improve=6.951438, (0 missing)
 DURTEMEL < 9.5 to the left, improve=3.216216, (0 missing)
 NUMNOEND < 40 to the right, improve=2.842883, (0 missing)
 NUMMEEND < 1.5 to the right, improve=2.768099, (0 missing)

Surrogate splits:

LENTEMEL < 15.5 to the left, agree=1.000, adj=1.000, (0 split)
 NUMNOEND < 58.5 to the right, agree=0.861, adj=0.357, (0 split)
 NUMMEEND < 1.5 to the right, agree=0.861, adj=0.357, (0 split)
 POCOMEL < 1.5 to the left, agree=0.788, adj=0.018, (0 split)

Node number 6: 56 observations, complexity param=0.01870748

predicted class=0 expected loss=0.4821429

class counts: 29 27

probabilities: 0.518 0.482

left son=12 (37 obs) right son=13 (19 obs)

Primary splits:

NUMNOEND < 35 to the right, improve=3.731000, (0 missing)
 ZMUSICAC < 0.3121117 to the left, improve=3.440476, (0 missing)
 SECSSEND < 37.75 to the left, improve=2.974844, (0 missing)
 DURTEMEL < 10.5 to the left, improve=2.974844, (0 missing)
 ZMUSICCO < 0.484234 to the left, improve=1.155590, (0 missing)

Surrogate splits:

NUMMEEND < 1.5 to the right, agree=0.696, adj=0.105, (0 split)
 SECSSEND < 39 to the right, agree=0.696, adj=0.105, (0 split)
 DURTEMEL < 13.5 to the right, agree=0.696, adj=0.105, (0 split)
 LENTEMEL < 13.5 to the right, agree=0.696, adj=0.105, (0 split)
 NOTEREST < 0.5 to the right, agree=0.696, adj=0.105, (0 split)

Node number 7: 203 observations

predicted class=1 expected loss=0.2364532

class counts: 48 155

probabilities: 0.236 0.764

Node number 12: 37 observations

predicted class=0 expected loss=0.3513514

class counts: 24 13

probabilities: 0.649 0.351

Node number 13: 19 observations

predicted class=1 expected loss=0.2631579

class counts: 5 14

probabilities: 0.263 0.737

2. Stufe zweistufiger Klassifizierungsbaum:

Baumzusammenfassung:

```
> tree1class_2step2
```

```
n= 665
```

```
node), split, n, loss, yval, (yprob)
```

```
* denotes terminal node
```

```
1) root 665 301 1 (0.54736842 0.26766917 0.18496241)
2) ZMUSICCO>=0.338823 191 87 1 (0.54450262 0.38743455 0.06806283)
4) ZMUSICCO>=0.7213568 97 32 1 (0.67010309 0.27835052 0.05154639)
8) SIMILAR>=0.4554 83 23 1 (0.72289157 0.21686747 0.06024096) *
9) SIMILAR< 0.4554 14 5 2 (0.35714286 0.64285714 0.00000000) *
5) ZMUSICCO< 0.7213568 94 47 2 (0.41489362 0.50000000 0.08510638)
10) ZMUSICCO< 0.50904 71 35 1 (0.50704225 0.40845070 0.08450704)
20) LENTEMEL< 25.5 42 18 2 (0.42857143 0.57142857 0.00000000) *
21) LENTEMEL>=25.5 29 11 1 (0.62068966 0.17241379 0.20689655) *
11) ZMUSICCO>=0.50904 23 5 2 (0.13043478 0.78260870 0.08695652) *
3) ZMUSICCO< 0.338823 474 214 1 (0.54852321 0.21940928 0.23206751)
6) ZMUSICCO< 0.1411808 425 177 1 (0.58352941 0.22588235 0.19058824) *
7) ZMUSICCO>=0.1411808 49 20 3 (0.24489796 0.16326531 0.59183673) *
```

2. Stufe zweistufiger Klassifizierungsbaum: Vollständige Beschreibung:

```
rpart(formula = CONFIDEN ~ IDPRE + SIMILAR + MANRHYTH + MANINTER +
  MANCONT + NUMNOEND + NUMMEEND + SECSSEND + DURTEMEL +
  LENTEMEL +
  POCOMEL + JOINT52 + MANSECEN + MANOTEPH + MANPHMEL +
  NOTEREST +
  HARMFUNC + CHANGERA + ZMUSICCO + ZMUSICAC, data = treedata_2step,
  method = "class", control = list(maxsurrogate = 18, cp = 0.01))
n= 665
```

```
CP nsplit rel error xerror xstd
1 0.02823920 0 1.0000000 1.0000000 0.04264387
```

2 0.02657807	2 0.9435216 1.0099668 0.04267889
3 0.02325581	3 0.9169435 0.9700997 0.04251758
4 0.01993355	4 0.8936877 0.9568106 0.04245116
5 0.01328904	5 0.8737542 0.9169435 0.04221356
6 0.01000000	6 0.8604651 0.9335548 0.04231959

Node number 1: 665 observations, complexity param=0.0282392

predicted class=1 expected loss=0.4526316

class counts: 364 178 123

probabilities: 0.547 0.268 0.185

left son=2 (191 obs) right son=3 (474 obs)

Primary splits:

ZMUSICCO < 0.338823 to the right, improve=7.507677, (0 missing)

ZMUSICAC < 0.3710032 to the left, improve=4.764074, (0 missing)

DURTEMEL < 19.5 to the left, improve=3.337638, (0 missing)

MANPHMEL < 1.25 to the left, improve=2.546594, (223 missing)

SIMILAR < 0.41425 to the right, improve=2.373786, (0 missing)

Surrogate splits:

ZMUSICAC < 1.132064 to the right, agree=0.749, adj=0.126, (0 split)

Node number 2: 191 observations, complexity param=0.02657807

predicted class=1 expected loss=0.4554974

class counts: 104 74 13

probabilities: 0.545 0.387 0.068

left son=4 (97 obs) right son=5 (94 obs)

Primary splits:

ZMUSICCO < 0.7213568 to the right, improve=5.508353, (0 missing)

SIMILAR < 0.4554 to the right, improve=3.735017, (0 missing)

JOINT52 < 0.6886 to the right, improve=3.300468, (0 missing)

ZMUSICAC < -0.2985999 to the left, improve=3.021573, (0 missing)

NUMNOEND < 36.5 to the right, improve=2.335796, (0 missing)

Surrogate splits:

ZMUSICAC < -0.2985999 to the left, agree=0.623, adj=0.234, (0 split)

LENTEMEL < 18.5 to the right, agree=0.534, adj=0.053, (0 split)

SECSSEND < 26.375 to the left, agree=0.529, adj=0.043, (0 split)

NOTEREST < 0.5 to the right, agree=0.529, adj=0.043, (0 split)

SIMILAR < 0.41425 to the right, agree=0.524, adj=0.032, (0 split)

NUMNOEND < 38.5 to the left, agree=0.518, adj=0.021, (0 split)

NUMMEEND < 1.5 to the left, agree=0.518, adj=0.021, (0 split)

DURTEMEL < 8.5 to the right, agree=0.518, adj=0.021, (0 split)

JOINT52 < 0.85295 to the left, agree=0.518, adj=0.021, (0 split)

MANRHYTH < 0.5 to the left, agree=0.513, adj=0.011, (0 split)

MANCONT < 0.5 to the right, agree=0.513, adj=0.011, (0 split)

POCOMEL < 1.5 to the right, agree=0.513, adj=0.011, (0 split)

Node number 3: 474 observations, complexity param=0.0282392

predicted class=1 expected loss=0.4514768

class counts: 260 104 110

probabilities: 0.549 0.219 0.232

left son=6 (425 obs) right son=7 (49 obs)

Primary splits:

ZMUSICCO < 0.1411808 to the left, improve=12.283790, (0 missing)
 ZMUSICAC < 0.2085403 to the left, improve= 6.945731, (0 missing)
 DURTEMEL < 19.5 to the left, improve= 3.176803, (0 missing)
 MANRHYTH < 0.5 to the right, improve= 1.717081, (0 missing)
 NUMNOEND < 84.5 to the left, improve= 1.616857, (0 missing)

Node number 4: 97 observations, complexity param=0.01328904

predicted class=1 expected loss=0.3298969

class counts: 65 27 5

probabilities: 0.670 0.278 0.052

left son=8 (83 obs) right son=9 (14 obs)

Primary splits:

SIMILAR < 0.4554 to the right, improve=3.819845, (0 missing)
 JOINT52 < 0.6886 to the right, improve=2.912171, (0 missing)
 MANOTEPH < 6.75 to the right, improve=1.984615, (32 missing)
 ZMUSICAC < 1.215894 to the left, improve=1.441793, (0 missing)
 DURTEMEL < 12.5 to the left, improve=1.409650, (0 missing)

Surrogate splits:

DURTEMEL < 9.5 to the right, agree=0.897, adj=0.286, (0 split)

Node number 5: 94 observations, complexity param=0.02325581

predicted class=2 expected loss=0.5

class counts: 39 47 8

probabilities: 0.415 0.500 0.085

left son=10 (71 obs) right son=11 (23 obs)

Primary splits:

ZMUSICCO < 0.50904 to the left, improve=4.896106, (0 missing)
 DURTEMEL < 14.5 to the right, improve=1.753423, (0 missing)
 LENTEMEL < 25.5 to the left, improve=1.662734, (0 missing)
 JOINT52 < 0.61805 to the right, improve=1.420441, (0 missing)
 ZMUSICAC < -0.07323501 to the left, improve=1.343060, (0 missing)

Node number 6: 425 observations

predicted class=1 expected loss=0.4164706

class counts: 248 96 81

probabilities: 0.584 0.226 0.191

Node number 7: 49 observations

predicted class=3 expected loss=0.4081633

class counts: 12 8 29

probabilities: 0.245 0.163 0.592

Node number 8: 83 observations

predicted class=1 expected loss=0.2771084

class counts: 60 18 5

probabilities: 0.723 0.217 0.060

Node number 9: 14 observations

predicted class=2 expected loss=0.3571429

class counts: 5 9 0

probabilities: 0.357 0.643 0.000

Node number 10: 71 observations, complexity param=0.01993355

predicted class=1 expected loss=0.4929577

class counts: 36 29 6

probabilities: 0.507 0.408 0.085

left son=20 (42 obs) right son=21 (29 obs)

Primary splits:

LENTEMEL < 25.5 to the left, improve=4.098800, (0 missing)

JOINT52 < 0.61805 to the right, improve=2.130730, (0 missing)

DURTEMEL < 14.5 to the left, improve=1.821817, (0 missing)

MANPHMEL < 3.25 to the left, improve=1.601717, (24 missing)

NUMNOEND < 36.5 to the right, improve=1.461157, (0 missing)

Surrogate splits:

SECSSEND < 20.05 to the right, agree=0.732, adj=0.345, (0 split)

DURTEMEL < 14.5 to the left, agree=0.704, adj=0.276, (0 split)

MANINTER < 0.5 to the left, agree=0.690, adj=0.241, (0 split)

NUMNOEND < 11 to the right, agree=0.690, adj=0.241, (0 split)

NUMMEEND < 0.55 to the right, agree=0.690, adj=0.241, (0 split)

POCOMEL < 1.5 to the left, agree=0.690, adj=0.241, (0 split)

NOTEREST < 0.5 to the left, agree=0.648, adj=0.138, (0 split)

SIMILAR < 0.7235 to the right, agree=0.634, adj=0.103, (0 split)

JOINT52 < 0.6886 to the left, agree=0.620, adj=0.069, (0 split)

Node number 11: 23 observations

predicted class=2 expected loss=0.2173913

class counts: 3 18 2

probabilities: 0.130 0.783 0.087

Node number 20: 42 observations

predicted class=2 expected loss=0.4285714

class counts: 18 24 0

probabilities: 0.429 0.571 0.000

Node number 21: 29 observations

predicted class=1 expected loss=0.3793103

class counts: 18 5 6

probabilities: 0.621 0.172 0.207

Nach pruning des Baums der zweiten Stufe:

Baumzusammenfassung:

> ptcl_2step2

n= 665

node), split, n, loss, yval, (yprob)

* denotes terminal node

1) root 665 301 1 (0.54736842 0.26766917 0.18496241)

2) ZMUSICCO>=0.338823 191 87 1 (0.54450262 0.38743455 0.06806283)

4) ZMUSICCO>=0.7213568 97 32 1 (0.67010309 0.27835052 0.05154639) *

- 5) ZMUSICCO < 0.7213568 94 47 2 (0.41489362 0.50000000 0.08510638)
 10) ZMUSICCO < 0.50904 71 35 1 (0.50704225 0.40845070 0.08450704) *
 11) ZMUSICCO >= 0.50904 23 5 2 (0.13043478 0.78260870 0.08695652) *
 3) ZMUSICCO < 0.338823 474 214 1 (0.54852321 0.21940928 0.23206751)
 6) ZMUSICCO < 0.1411808 425 177 1 (0.58352941 0.22588235 0.19058824) *
 7) ZMUSICCO >= 0.1411808 49 20 3 (0.24489796 0.16326531 0.59183673) *

A-5.6.4 Diskussion der statistischen Modellierung

Residuen des Regressionsbaummodells:

