

ADAPTIVE MULTIMODALE EXPLORATION VON MUSIKSAMMLUNGEN

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften
der RWTH Aachen University zur
Erlangung des akademischen Grades eines Doktors der
Naturwissenschaften genehmigte Dissertation

vorgelegt von
Diplom-Informatiker

DOMINIK LÜBBERS
aus Osnabrück

Berichter:

Universitätsprofessor Dr. rer. pol. Matthias Jarke
Universitätsprofessor Dr. rer. nat. Thomas Seidl

Tag der mündlichen Prüfung: 8. Dezember 2009

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online verfügbar.

Zusammenfassung

Aktuelle Musikportale und Systeme zur Verwaltung privater Musiksammlungen konzentrieren sich auf die Unterstützung einer gezielten Suche nach Stücken (z.B. anhand von beschreibenden Metadaten wie Titel, Künstler, etc.). Empirische Untersuchungen haben jedoch gezeigt, dass das typische Kundenverhalten in klassischen Musikläden viel stärker durch ein ungerichtetes „Stöbern“ im Angebot ohne ein klar artikulierbares Informationsbedürfnis gekennzeichnet ist.

Um dieses Zugriffsmuster besser zu bedienen, beschreibt diese Dissertation die Konzeption, Implementierung und Evaluation einer Computerspiel-artigen *Explorationsumgebung für Musiksammlungen*, die dem Benutzer die Freiheit zur selbständigen Erkundung einer nachvollziehbar strukturierten Stück-Kollektion gibt.

Grundlage einer solchen Strukturierung ist ein quantifizierbarer Ähnlichkeitsbegriff für Musiktitel. Im ersten Teil der Arbeit werden daher inhaltsbasierte, Experten-basierte, kollaborative und textbasierte Ansätze zur Distanzberechnung diskutiert und bewertend gegenübergestellt. Zur Modellierung des Facettenreichtums musikalischer Ähnlichkeit wird eine Auswahl dieser Maße, die unterschiedliche musikalische Aspekte betrachten, zu einer zusammenfassenden Distanzfunktion kombiniert.

Die Generierung selbstorganisierender Merkmalskarten erlaubt die Topologie-erhaltende Anordnung von Stücken auf einer zweidimensionalen Landkarte, wobei zueinander ähnliche Titel in räumlicher Nähe positioniert werden. Ferner ermöglicht sie die Identifikation von Cluster-grenzen, die homogene Gruppen ähnlicher Stücke voneinander separieren. Diese Informationen lassen sich zum Aufbau einer dreidimensionalen virtuellen Landschaft verwenden, in der der Benutzer frei navigieren kann.

Ein zentraler Beitrag dieser Arbeit besteht in der Ergänzung dieser visuellen Darstellung um die permanente direktionale Wiedergabe der positionierten Stücke, die eine Medienbruch-freie Kommunikation des eigentlichen Explorationsgegenstandes ermöglicht und so zu einer informierteren Bewegung des Benutzers in der Landschaft führt. Zur Vermeidung einer akustischen Reizüberflutung wurden Konzepte zur Wahrnehmungs-Fokussierung und gezielten Wiedergabeauswahl erarbeitet.

Ferner gestattet die entwickelte Umgebung nicht nur die Anpassung der Umgebung *durch* den Benutzer (z.B. durch die Verschiebung von Stücken oder Umformungen des Terrains), sondern realisiert ebenso eine Anpassung *an* den Benutzer: Auf der Grundlage der Interaktion des Anwenders mit dem System wird eine Neu-Gewichtung der Komponenten des Ähnlichkeitsmaßes gelernt, die die Strukturierungsvorgaben des Benutzers bestmöglich erklären. Die Entwicklung dieses adaptiven Distanzmaßes zur Modellierung eines personalisierten Ähnlichkeitsbegriffes stellt einen weiteren Kernbeitrag dieser Dissertation dar.

In einer abschließenden Evaluation wurde die Bedeutung der *multimodalen* Darstellung für die Orientierung des Benutzers in der virtuellen Umgebung sowie die Leistungsfähigkeit der *adaptiven* Distanzfunktion nachgewiesen.

Abstract

Current online music portals and systems for the management of private music collections mainly concentrate on the support for directed search for music pieces (e.g. on the basis of describing metadata like title, artist, etc.). However, empirical studies have shown that the typical behaviour of customers in classic music retail stores is much more characterized by an undirected browsing in the offered catalogue without being able to articulate a clear information demand.

To improve support for this access paradigm, this thesis describes the conceptualization, implementation and evaluation of a computer game-like *exploration environment for music collections*, which allows its users to freely and autonomously explore a comprehensibly structured music library.

Such a structure is based on a quantifiable notion of similarity between songs. Therefore, in the first part of this thesis content-based, expert-based, collaborative and text-based approaches for distance calculation are discussed and assessed. To model the multiple facets of music similarity, a subset of these measures, which cover different aspects of music, are combined into a summarizing distance function.

The generation of self-organizing maps allows the topology-preserving arrangement of pieces on a two-dimensional map, on which similar songs are placed close to each other. Furthermore, it allows the detection of cluster boundaries, that separate homogeneous groups of similar pieces. This information can be used to build a three-dimensional virtual landscape, which the user can freely navigate in.

A main contribution of this thesis is the supplementation of this visual presentation by permanent spatialized acoustic playback of the placed pieces, which makes a media break-free communication of the actual exploration subject possible and thereby allows for a more informed navigation of the user in the landscape. To prevent users' acoustic overstimulation we developed and implemented concepts for perception focussing and intelligent selection of songs for playback.

Besides the development of possibilities to customize the environment (e.g. by moving songs or changing the height profile of the terrain), we also let the environment adapt to the user automatically: On the basis of his interaction with the system, a re-weighting of the individual similarity measure components is learned, which explains the structure provided by the user best. The development of this adaptive distance measure which allows to model a personalized similarity notion can be regarded as a further main contribution of this thesis.

A concluding evaluation shows the importance of the developed *multimodal* presentation for the user's orientation in the virtual environment and proves the performance of the *adaptive* distance function.

Danksagung

Die Anfertigung dieser Dissertation wäre ohne die vielfältige Unterstützung zahlreicher Personen nicht möglich gewesen. Im Bewusstsein, nicht alle aufführen zu können, möchte ich insbesondere folgenden Menschen meinen Dank aussprechen:

- meinem Doktorvater Herrn Prof. Jarke für die vertrauensvolle Freiheit, selbständig ein spannendes Themenfeld wählen zu können, seine konstruktive Unterstützung bei der Erstellung dieser Arbeit sowie seine fürsorgliche Betreuung, insbesondere seinen Einsatz zur Ermöglichung eines in vieler Hinsicht aufregenden Auslandsaufenthaltes an der GUTech in Muscat,
- Herrn Prof. Seidl für seine unkomplizierte Bereitschaft zur prompten Übernahme des Zweitgutachtens,
- meinen Diplomanden und Master-Studenten Jan Neis, Torben Keller und Qian Yang für ihre konstruktiven und hilfreichen Beiträge,
- meinen Kolleginnen und Kollegen vom i5 für fachlich anspruchsvolle Gespräche und entspanntes Beisammensein,
- meinen Kolleginnen und Kollegen an der GUTech für die zahlreichen Diskussionen und Gespräche während der gemeinsamen Zeit in den anfänglichen Wirren der Gründungsphase der Universität,
- Dominik Schmitz, mit dem ich während eines Großteils meiner Zeit am i5 das Büro (die „Dominikanische Republik“) teilen durfte, für erhellende und inspirierende Gespräche, aufmunternde Worte in so einigen Motivationstiefen und ein offenes Ohr in allen Lebenslagen,
- meiner Familie sowie meiner Freundin Nina Hilberink für ihre allgegenwärtige und vielfältige Unterstützung.

Danke.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Von der CD zum Download	1
1.1.1	Tonträger-Verkäufe und Tauschbörsen	1
1.1.2	Reaktion der Musikindustrie	2
1.1.3	Vorteile digitaler Musikdistribution	4
1.2	Exploration von Musiksammlungen	8
1.2.1	Gezielte Suche vs. explorativer Zugriff	8
1.2.2	Explorations-Unterstützung aktueller Musikportale	8
1.2.3	Kritik aktueller Explorations-Unterstützung	9
1.2.4	Nutzer-adaptive Exploration	10
1.2.5	Zielsetzung	10
1.3	Repräsentationsformen von Musik	11
1.3.1	Musik als Noten	11
1.3.2	Musik als Steuersignal	15
1.3.3	Musik als Wellenform	16
1.3.4	Wechsel des Abstraktionsniveaus	19
1.4	Prototyp soniXplorer	25
1.5	Aufbau der Arbeit	28
1.5.1	Struktur	28
1.5.2	Kern-Beiträge der Arbeit	29
2	Ähnlichkeitsmaße	31
2.1	Inhaltsbasierte Ähnlichkeit	33
2.1.1	Klang-basierte Merkmale	33
2.1.2	Dynamik- und Rhythmus-basierte Merkmale	41
2.2	Experten-basierte Ähnlichkeit	46
2.2.1	AMG: Annotierter Musikkatalog	47
2.2.2	Music Genome Project (Pandora)	48

2.3	Kollaborative Ähnlichkeit	49
2.3.1	Kaufhistorien und Benutzerbeobachtung	50
2.3.2	Gemeinsames Auftreten in Playlists	51
2.3.3	Tagging	53
2.3.4	Diskussion	55
2.4	Textbasierte Ähnlichkeit	57
2.4.1	Semantische Analyse von Liedtexten	58
2.4.2	Textuelle Ähnlichkeit nach Whitman	58
2.4.3	Textuell-inhaltsbasierte Ähnlichkeit nach Knees	59
2.4.4	Berücksichtigung von Nutzer-Feedback	60
2.4.5	Diskussion	63
2.5	Zusammenfassende Bewertung	63
2.6	Kombination von Ähnlichkeitsmaßen	66
2.7	Inhaltsbasierte Simulation kollaborativer Ähnlichkeit	68
2.8	Ähnlichkeitsmaße im soniXplorer	72
2.8.1	Inhaltsbasierte Merkmale	72
2.8.2	Metadaten-basierte Merkmale	73
2.8.3	Kollaborative Merkmale	73
2.8.4	Fehlende Werte	76
3	Explorationsumgebung	79
3.1	Forschungsumfeld	79
3.1.1	Metadaten-basierte Visualisierung	80
3.1.2	Sonic Browser	81
3.1.3	Islands of Music	82
3.1.4	MusicMiner	83
3.1.5	Akustische Exploration	85
3.1.6	Dreidimensionale multimodale Explorationsumgebung	86
3.1.7	Zusammenfassung und Gegenüberstellung	89
3.2	Vektorraum-Repräsentation	91
3.3	Strukturierung mit selbstorganisierenden Merkmalskarten	96
3.3.1	Formalisierung	98
3.3.2	Lernverfahren	99
3.3.3	Karten-Topologie	100
3.3.4	Höhenprofil-Berechnung	103
3.4	Multimodale Präsentation	104
3.4.1	Visuelle Darstellung	107
3.4.2	Akustische Darstellung	111
3.4.3	Navigation	120

4	Interaktion und Nutzer-Adaption	123
4.1	Interaktion	123
4.1.1	Anpassung der Stück-Darstellung	124
4.1.2	Orientierungspunkte	125
4.1.3	Anpassung des Terrains	127
4.2	Adaption	128
4.2.1	Verwandte Arbeiten	129
4.2.2	Anpassung des Ähnlichkeitsmaßes	138
4.2.3	Anpassung der Vektor-Repräsentation	142
4.2.4	Anpassung der selbstorganisierenden Karte	143
5	Evaluation	145
5.1	Fragestellungen und Untersuchungshypothesen	145
5.2	Evaluations-Design	146
5.2.1	Quantitative Auswertung	146
5.2.2	Subjektive Auswertung	149
5.3	Ergebnisse und Auswertung	150
5.3.1	Versuchsgruppe	150
5.3.2	Quantitative Auswertung	155
5.3.3	Subjektive Bewertung	159
5.3.4	Strukturierte Kollektionsdarstellung	159
5.3.5	Visuelle Darstellung	160
5.3.6	Akustische Darstellung	161
5.3.7	Hinweisschilder	162
5.3.8	Allgemeine Benutzerfreundlichkeit	163
6	Zusammenfassung und Ausblick	165
6.1	Zusammenfassung	165
6.2	Zukünftige Arbeiten	167
6.2.1	Skalierbarkeit	167
6.2.2	Erweiterungen der akustischen Titel-Wiedergabe	169
6.2.3	(Semi-)Automatische Annotation von Kartenregionen	170
6.2.4	Alternative Ähnlichkeits-Kombinationsmodelle	170
6.2.5	Umfangreiche Evaluationsstudien	170
A	Verwendete Musik-Kollektionen	171
A.1	Demonstrations-Kollektion	171
A.2	Kollektion strukturierte Darstellung	173
A.3	Kollektionen akustische Wiedergabe	176
A.4	Kollektion Benutzeradaption	179

B Evaluations-Fragebogen	181
Literaturverzeichnis	186
Abbildungsverzeichnis	205

Kapitel 1

Einleitung

1.1 Von der CD zum Download

1.1.1 Tonträger-Verkäufe und Tauschbörsen

Seit mehreren Jahren ist die Musikindustrie mit einem zum Teil deutlichen Einnahmen-Rückgang aus den Verkäufen physikalischer Tonträger konfrontiert. Abbildung 1.1 fasst die vom IFPI (International Federation of the Phonography Industry) veröffentlichten Einzelstatistiken über die Jahre 1995 bis 2007 zusammen und zeigt einen klaren Abwärtstrend seit dem Jahr 2000.

Eine Reihe ökonometrischer Studien hat untersucht, inwiefern dieser Wendepunkt nicht nur zufällig mit der Etablierung illegaler Filsharing-Angebote wie dem im Juni 1999 gestarteten Napster-System zusammenfällt [Bla04, Hon04a, Hon04b, Mic04, Zen05, OS07, PW04, RW06].

Gemäß Zentner sinkt die Wahrscheinlichkeit, dass ein Kunde Musik auf physikalischem Träger erwirbt, mit der intensiven Nutzung von Peer-To-Peer-Tauschbörsen um 30% [Zen06]. Michel führt zumindest 5% des Verkaufsrückganges auf Filesharing zurück [Mic04], Hong beziffert diesen Anteil auf 7,6% [Hon04a].

Die angewandten Methoden, mit denen der Einfluss von Tauschbörsen auf das Kaufverhalten von Kunden quantifiziert wird, sind vielfältig. Zumeist wird versucht, eine Korrelation zwischen beobachteten Verkaufszahlen und Messgrößen, die die Nutzungsintensität von Tauschbörsen widerspiegeln, zu belegen. So vergleicht Zentner [Zen05] beispielsweise landesweite Verkaufszahlen (nach IFPI-Statistiken) mit lokaler Breitband-Internet-Nutzung (gemäß Daten der International Telecommunication Union (ITU)) und kann einen signifikanten Zusammenhang nachweisen.

Eine viel zitierte Studie, die einen klaren Einfluss von Tauschbörsen-Aktivität auf Tonträger-Verkaufszahlen verneint, ist die empirische Analyse von Oberholzer und Strumpf [OS07]. Ihrer Meinung nach wird die Ersetzung eines CD-Kaufs durch illegalen Download von einer verstärkten Exposition des Benutzers mit ihm bislang nicht bekannter Musik begleitet, die er beim Durchstöbern der Ordner anderer Nutzer kennenlernt und z.B. in Chatrooms mit ihnen diskutieren kann. Dieser implizite Werbeeffect von Tauschportalen motiviere zum CD-Kauf und habe einen positiven kompensatorischen Effekt auf Tonträger-Verkaufszahlen.

Oberholzer und Strumpf kommen durch eine Analyse von Server-Logfiles der Tauschbörse OpenNap und Album-Verkaufszahlen gemäß Nielsen Soundscan-Statistiken zu dem Schluss,

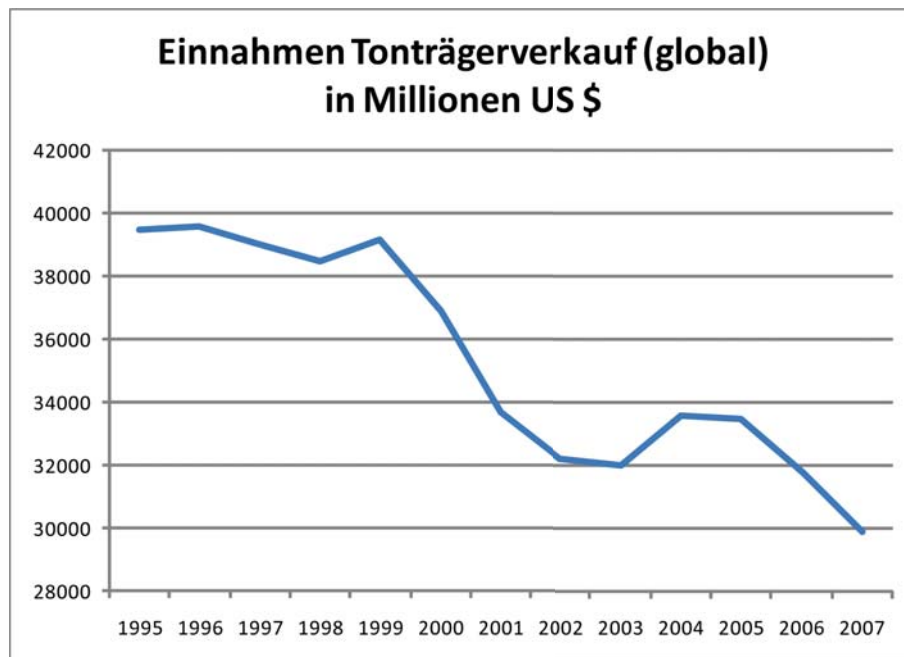


Abbildung 1.1: Einnahmen aus dem Verkauf von physikalischen Tonträgern weltweit.
Quelle: [IFP07]

dass sich kein signifikanter Einfluss von Tauschbörsen-Nutzung auf CD-Verkaufszahlen nachweisen lasse und der deutliche Rückgang durch andere Faktoren erklärt werden müsse.

Die trotz Download-Alternativen große Bedeutung physikalischer und damit *greifbarer* Tonträger wird in zwei Studien von Cunningham et al. herausgearbeitet, die das Suchverhalten von Kunden in Musikhandlungen und Musikbibliotheken [CRB03] sowie die Organisation persönlicher Musiksammlungen [CJJ04] analysieren. Neben der eigentlichen Audio-Information erleben die beobachteten Studien-Teilnehmer die „taktile Präsenz“ der physikalischen CD als wichtigen Teil ihres Kauf- und Besitz-Erlebnisses. Eine besondere Bedeutung komme zudem dem Album-Cover zu, das gemäß Cunningham et al. für viele Kunden als wichtiger (und mitunter kaufentscheidender) Teil des Produktes wahrgenommen werde. Dies spiegle sich insbesondere in dem Aufwand wieder, den eine Reihe von Studien-Teilnehmern in die Gestaltung von CD-Covern selbst zusammengestellter Sammlungen investierten.

1.1.2 Reaktion der Musikindustrie

Als Reaktion auf die rückläufigen Einnahmen aus dem Verkauf physikalischer Tonträger reagierte die internationale Musikindustrie zunächst protektionistisch. Auf juristischem Wege versuchte man, die Nutzung von Tauschbörsen unter Verweis auf Urheberrechts-Verletzungen einzudämmen. So klagte A&M-Records 2001 erfolgreich gegen den Tauschserver-Betreiber Napster und erzwang die Einstellung seines Betriebs.

Als Reaktion auf die Schließung entwickelten sich jedoch bald dezentrale Tausch-Protokolle, die die Server-Funktionalität auf die einzelnen Clients verteilen und das Netz somit robuster gegen Abschaltungen einzelner Peers machen.

Daraufhin verklagte die Musikindustrie einzelne hochaktive Peer-To-Peer-Nutzer. Zwischen 2003 und 2008 kam es so zu ungefähr 35000 Einzelanklagen wegen Verletzung des Urheberrechts [MS08b].

Ende 2008 verkündete die Interessensvertretung der US-amerikanischen Musikindustrie RIAA jedoch einen Strategiewechsel zur Eindämmung von illegalen Tauschaktivitäten: In Zukunft werde man auf Einzelklagen verzichten und statt dessen versuchen, mit verschiedenen Internet-Service-Providern an einer Reduktion des illegalen Musik-Download-Volumens zu arbeiten.

Detaillierte Informationen zur Begründung dieses neuen Vorgehens sind nicht veröffentlicht; es lässt sich aber annehmen, dass das zunehmend schlechte Bild der Musikindustrie in der Öffentlichkeit sowie die immer wieder in Frage gestellte Wirksamkeit isolierter Einzelklagen ausschlaggebend waren.

Seit ungefähr 2004 hat sich zudem eine weitere Trendwende der Musikindustrie im Umgang mit online vertriebener Download-Musik vollzogen: Versuchte man anfangs, die Distribution von Musik im Internet weitestmöglich einzudämmen, so engagiert man sich nun selbst mit dem Betrieb kostenpflichtiger Online-Musikportale und stellt somit konstruktiv eine legale Alternative zu illegalen Tauschbörsen bereit. Zur Verhinderung einer unkontrollierten Ausbreitung von zum Download angebotener Musik werden die meisten Stücke mit Methoden des Digital Right Managements (DRM) gesichert, die die Anzahl möglicher Kopien beschränken und mitunter eine Identifizierung des Kunden ermöglichen, der die Datei ursprünglich erworben hat. In den letzten Jahren ist aufgrund der immer wieder auftretenden technischen Probleme mit DRM-geschützten Material und des Widerstandes von Benutzern gegen die Einschränkung des Umgangs mit rechtmäßig erworbener Musik jedoch zunehmend ein Trend zur Verbreitung von DRM-freien Stücken zu erkennen (Musicload.de vertreibt seit dem 1. März 2009 alle Stücke DRM-frei, iTunes hat eine Umstellung seines gesamten Kataloges auf DRM-freie Titel bis zum 2. Quartal 2009 angekündigt).

Abbildung 1.2 illustriert die zunehmende Akzeptanz des Angebotes legaler Download-Musikportale durch den Kunden und die steigende Bedeutung der Einnahmen aus dem Verkauf von online vertriebener Musik. In den Statistiken der IFPI wird stets zwischen dem Vertrieb von *Musik auf physischem Tonträger* (wie z.B. CD, Vinyl-Platte, etc.) und dem Angebot *digitaler Musik* unterschieden. Letzteres meint eigentlich die Bereitstellung komprimierter Musikdateien (z.B. im MP3-Format), die eine effiziente Übertragung über das Internet und eine kompakte Speicherung beim Kunden ermöglichen. Die Bezeichnung *digitale Musik* ist damit etwas irreführend, da das Musiksinal auf einer CD ebenfalls in digitalisierter Form vorliegt.

In den letzten Jahren haben sich im Wesentlichen drei Geschäftsmodelle zur legalen Distribution von digitaler Musik im Internet herausgebildet:

1. A-la-carte-Musik-Portale

A-la-carte-Musik-Portale stellen dem Benutzer einen (mehr oder minder umfangreichen) Katalog verfügbarer Musik zum Download zur Verfügung. Meist kann der Kunde zwischen dem Erwerb einzelner Stücke oder dem Kauf aller Tracks eines Albums wählen. Letzteres ist in der Regel jedoch kaum günstiger als der Erwerb des physischen Tonträgers.

Die bedeutendsten internationalen Internet-Portale dieses Typs sind der iTunes-Musicstore von Apple und das Download-Angebot von Amazon.com. In Deutschland ist auch das von T-Online betriebene Portal Musicload.de etabliert.

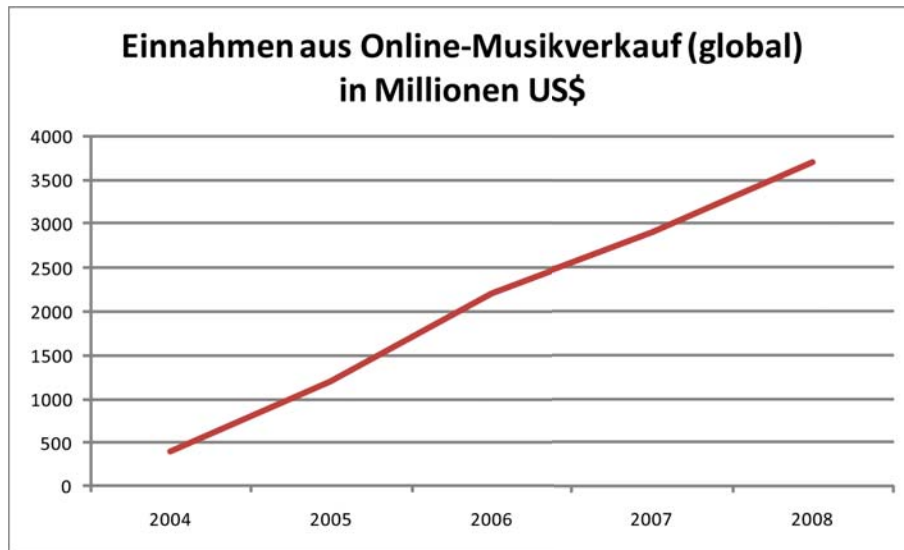


Abbildung 1.2: Einnahmen aus dem Online-Verkauf digitaler Musik weltweit. Quelle: [IFP09]

2. Abo-Dienste

Gegen eine meist monatliche Gebühr erlauben Abo-Portale ihren Kunden die unbegrenzte Wiedergabe von Stücken aus dem Katalog des Anbieters. Im Gegensatz zu den A-la-carte-Musik-Portalen ist ein Download der ausgewählten Tracks nicht möglich; anstatt dessen wird die Musik direkt vom Portal-Betreiber zum Rechner des Kunden gestreamt. Somit wird der Aufbau einer privaten Musikkollektion beim Kunden zugunsten einer Bindung an den Abo-Dienst verhindert.

Die bekanntesten Musik-Portale mit einem umfangreichen Abo-Service sind Napster und Rhapsody. Auch Musicload.de in Deutschland bietet eine sog. Musik-Flatrate an.

3. Werbe-finanzierte Dienste

Eine Reihe von Internet-Portalen refinanziert die Urheberrechtsabgaben für die legale Bereitstellung gestreamter Musik durch die Einblendung von Werbung. Prominentester Vertreter dieses Geschäftsmodells ist YouTube, das eine Fülle von Musikvideos anbietet und mitunter als „digitaler Nachfolger des Musikfernsehsenders MTV“ bezeichnet wird [LeC09].

1.1.3 Vorteile digitaler Musikdistribution

Obige Zahlen legen die Prognose eines weiteren deutlichen Wachstums der Online-Distribution im Vergleich zum Vertrieb von Musik auf physikalischem Tonträger nahe. Im folgenden werden einige der anzunehmenden Gründe für diese Entwicklung zusammengetragen und beleuchtet. Hierbei lässt sich zwischen Vorteilen durch die Bereitstellung von Musikstücken als komprimierte Dateien und positiven Effekten, die sich aus der Präsentation von Musik auf Online-Musik-Portalen ergeben, unterscheiden.

Musik als komprimierte Datei

Grundlage der Geschäftsmodelle zur elektronischen Distribution von Musik im Internet ist die Verfügbarkeit von Kompressionsverfahren, die die immense Datenmenge digitalisierter Musiksignale mit vernachlässigbarem hörbarem Qualitätsverlust in kompakten Dateien enkodieren. Prominenter Vertreter dieser Gruppe ist sicherlich der MP3-Codec (eigentlich MPEG-1 Audio Layer 3) [Bra99], der insbesondere eine effiziente Dekodierung des Datenstromes auf dem Wiedergabegerät ermöglicht. Grundlage der verlustbehafteten Kompression ist die Ausnutzung physio- und psychoakustischer Forschungsergebnisse über das menschliche Gehör, wie z.B. der begrenzten Frequenzauflösung (erklärbar durch die Anordnung der Sinneszellen in der Gehörschnecke) oder von Maskierungseffekten, die die Überdeckung von benachbarten Frequenzen durch einen dominanten Ton beschreiben. Der MP3-Codec erlaubt verschiedene Qualitätsstufen. Ab einer Datenrate von 160 kbps können auch geübte Hörer unter konzentrierten Bedingungen keinen hörbaren Unterschied zwischen dem digitalen Original und dem kodierten Signal feststellen [Mey00]. Bei dieser Datenrate ergibt sich gegenüber der CD (Stereo-Signal, 16-Bit Samples, Abtastrate von 44100 Hz) ein Kompressionsfaktor von $\frac{160 \text{ kbps}}{1411 \text{ kbps}} \approx 11\%$.

Darüber hinaus existiert eine Fülle weiterer Kompressionsverfahren (AAC, Ogg Vorbis, WMA, ...), die mitunter zusätzlich Funktionen des Digital Rights Managements (DRM) implementieren.

Die Kompaktifizierung von digitaler Musik in Form komprimierter Dateien bietet dem Kunden folgende Vorteile:

- *Kollektionsvolumen*

Die geringe Dateigröße komprimierter Musikstücke erlaubt zusammen mit der steigenden Speicherkapazität von Festplatten und in mobilen Abspielgeräten verwendeten Festspeichern sowie der wachsenden Internet-Bandbreite den Aufbau privater Musikkollektionen, deren Größe beim Erwerb der entsprechenden physikalischen Tonträger nicht realisierbar ist.

Zum großen Volumen privater Musiksammlungen trägt zudem die unkomplizierte (meist illegale) Kopierbarkeit von nicht DRM-geschützter Musik bei. Viele Nutzer „teilen“ ihre Musikkollektion mit Freunden und Bekannten und erweitern so ihre eigene Sammlung.

- *Lieferzeit*

Wenn sich ein Kunde zum Kauf eines Musikstückes im Internet entscheidet, ersetzt ein schneller Download die ansonsten zeitraubende und Ressourcen-intensive Versendung des physikalischen Tonträgers. Das Produkt ist damit bereits kurz nach der Kaufentscheidung bereits zum Kunden ausgeliefert.

- *Kollektions-Mobilität*

Der geringe Speicherbedarf komprimierter Musikdateien und die hohe Speicherkapazität mobiler Abspielgeräte erlauben heute, große Teile der eigenen Musikkollektion permanent mit sich zu führen. Zusammen mit ausgereiften Synchronisationsmechanismen zwischen Abspielgerät und stationärem Kollektionsspeicher wird dadurch eine mit physikalischen Tonträgern nicht realisierbare Mobilität ermöglicht.

- *Stück-selektiver Einkauf*

Musikstücke können getrennt von ihrem Album-Kontext angeboten und erworben werden. Dies erlaubt dem Kunden einen selektiveren Zugang zu Musik, die er mag und für die er bereit ist, Geld zu bezahlen.

- *Preis*

Der Betrieb eines Online-Portals zum Vertrieb von komprimierter Musik erfordert (abgesehen von den einmalig zu treffenden Vereinbarungen mit den Rechteinhabern) vom Unternehmer einen deutlich geringen Einsatz finanzieller Ressourcen. Insbesondere entfallen kostspielige Investitionen in Lagerhaltung und Versandabteilung. Diese Einsparungen können an den Kunden weitergegeben werden und führen letztlich zu niedrigeren Preisen.

- *Angebotsbreite*

Die Aufnahme eines Albums in den Angebotskatalog eines Portal-Betreibers ist - nach Abklärung der Fragen zur Urheberrechtsvergütung - mit nur geringem technischen Aufwand verbunden. Dies ermöglicht dem Unternehmer den Aufbau eines weitaus größeren und vielfältigeren Angebotes als dies mit physikalischen Tonträgern möglich wäre. Die herabgesenkte Schwelle zur Aufnahme in einen Verkaufskatalog kommt vor allem unbekannten Musikschaaffenden zu Gute, die nicht so sehr im Fokus der Werbestrategien der Musikindustrie stehen.

Musikkauf in Online-Portalen

Im folgenden werden einige Faktoren zusammengetragen, die den wachsenden Erfolg von Online-Portalen zum Erwerb von Musik erklären. Diese Faktoren sind im Wesentlichen unabhängig von der Lieferform der angebotenen Musik und gelten somit in weiten Teilen auch für Einkaufs-Portale wie Amazon.com, die (möglicherweise ergänzend zu einem Download-Angebot) den Versand von Musik auf physikalischen Tonträger anbieten.

- *Verfügbarkeit*

Online-Portale sind nicht an Ladenöffnungszeiten gebunden und offerieren (bis auf kurze Wartungsintervalle) dem Kunden eine ständige Verfügbarkeit. Hinzu kommt die Ortsunabhängigkeit des Zugriffs auf das Portal: Kunden können von jedem Internet-Zugang aus im Angebotskatalog stöbern und Bestellungen aufgeben.

- *Umfeldinformationen*

Viele Online-Portale ergänzen ihr Musik-Angebot durch Hintergrund-Informationen zum Produkt, die vom Kunden nach Bedarf abgerufen (und mitunter auch ergänzt) werden können. So bietet Last.fm neben biographischen Informationen zum Künstler eine Übersicht über dessen anstehende Konzerte und Tourneen, AllMusicGuide.com verlinkt Künstler, die nach Meinung der Musikredakteure ähnlich zueinander sind. Fast alle Online-Portale ordnen die angebotene Musik zumindest einem oder mehreren Genres zu.

- *Vorhörfunktion*

Seit einigen Jahren kann der Kunde einen wachsenden Teil des Musikangebotes vor einem potentiellen Kauf vorhören. Meist ist diese Vorhörfunktion auf einen Ausschnitt des Gesamttitels beschränkt und/oder erfolgt in geringerer Klangqualität. Damit ermöglichen die Online-Portale dem Benutzer, seine Kaufentscheidung fundiert auf Grundlage eines „Tests“ des letztlichen Produktes zu treffen.

Die meisten Musikhandlungen räumen seit geraumer Zeit Ihren Kunden ein ähnliches Probehören an dezidierten Abhörstationen ein. Wie Cunningham et al. durch Beobachtung und Befragung von Probanden beim Einkauf in einem Musikladen herausfanden, überschreiten die meisten Kunden jedoch nicht die Hemmschwelle, einen Verkäufer anzusprechen [CRB03]. Diese Schwelle wird durch die anonyme Atmosphäre eines Online-Shops signifikant gesenkt.

- *Kollaborativer Austausch, Community-Funktionen*

Der Musik-Einkauf ist gemäß Cunningham ein nicht selten kollaborativer Prozess, bei dem sich der Kunde auf die Meinung seiner Umgebung stützt [CRB03]. Der Bedeutung der Erfahrung anderer für den persönlichen Einkauf wird in den meisten Online-Portalen durch das Angebot eines Bewertungs-Systems Rechnung getragen, in dem Kunden das von ihnen erworbene Produkt beurteilen können und diese Bewertung für spätere Interessenten sichtbar ist. Amazon.com erlaubt darüber hinaus das Verfassen von textuellen Nutzerrezensionen.

Andere Online-Portale räumen ihren Benutzern noch weitergehende Mitgestaltungsmöglichkeiten ein. So ermuntert Last.fm seine Kunden, verfügbare Kontextinformationen zu Künstlern oder Alben zu ändern oder zu ergänzen. Hierzu wird ein Wiki-ähnliches System bereitgestellt.

- *Empfehlungssystem*

Die große Benutzerzahl dominanter Online-Portale wie z.B. Amazon.com erlaubt durch Analyse von Kauf- und Navigations-Verhalten seiner Kunden die Identifikation von Nutzern, die ein zum aktuellen Benutzerprofil ähnliches Interessensspektrum aufweisen. Die Kaufhistorie dieser Nutzer kann schließlich zur Empfehlung von Produkten (und insbesondere Musik) verwendet werden, die dem Geschmack des aktuellen Kunden entsprechen (Details zur Generierung solcher kollaborativer Empfehlungen werden in Abschnitt 2.3 beschrieben).

Die Qualität kollaborativer Empfehlungen hängt maßgeblich von Umfang und Qualität der zugrundeliegenden Daten ab. Hier besitzt ein prominentes Online-Portal, das die Aktionen seiner Kunden genau protokolliert, entscheidende Vorteile gegenüber einem herkömmlichen meist lokalen Musikhändler und versucht so, das Fehlen eines persönlichen Einkaufsberaters zu kompensieren.

Die hohe Bedeutung von kollaborativen Empfehlungen für den Verkauf von Produkten im Internet lässt sich an der Tatsache ermessen, dass laut Stenzel und Kamps 60% - 70% der Seitenaufrufe bei Amazon.com durch das Folgen eines Item-to-Item-Empfehlungs-Links entstehen [SK05].

1.2 Exploration von Musiksammlungen

Die in den obigen Abschnitten beschriebene und begründete Tendenz zur Verbreitung von komprimierter Musik in Online-Portalen und das zu erwartende Wachstum persönlicher Musiksammlungen erfordern die Entwicklung geeigneter Zugriffsmechanismen zur Suche und Navigation in diesen Medienarchiven.

1.2.1 Gezielte Suche vs. explorativer Zugriff

Im Information Retrieval wird zwischen zwei Suchparadigmen unterschieden: Der *gezielten Suche* nach Objekten und dem explorativen Zugriff auf eine Objekt-Kollektion [BYRN99].

Voraussetzung der gezielten Suche ist ein klares Informationsbedürfnis beim Benutzer: Der Anwender hat eine präzise Vorstellung von dem zu findenden Objekt und kann dessen Charakteristika genau beschreiben. Einfaches Beispiel für diesen Suchtyp ist die Identifikation eines Musik-Albums durch Angabe des Künstlers und des Albumtitles. Da sich gezielte Suchanfragen in der Regel auf einfache Datenbankabfragen reduzieren lassen, unterstützen nahezu alle Musikportale diese Strategie und unterscheiden sich lediglich in dem Angebot an Metadaten, die zur Anfrageformulierung verwendet werden können.

Ein explorativer Zugriff kommt einem zunächst ungerichteten Durchstöbern oder Erkunden der Objektkollektion nahe. Der Benutzer greift ohne konkretes Informationsbedürfnis auf die Objektmenge zu – möglicherweise entwickelt sich sein Interesse für bestimmte Objekte erst während des Explorationsprozesses als Folge einer gewonnenen groben Orientierung. Damit gleicht die Exploration einem Browsing, das Marchionini und Shneiderman als „[...]an exploratory, information seeking strategy that depends upon serendipity“ [MS88] definieren. Die Aufgabe einer Explorationsumgebung besteht darin, dem Benutzer einen intuitiven Zugang zur Kollektion zu ermöglichen, der seinen Spürsinn (serendipity) für relevante Informationen unterstützt.

Cunningham et al. identifizieren den ungerichteten Zugriff als dominantes Zugriffsparadigma für Musikkollektionen [CRB03]. Die Beobachtung und Auswertung des Kaufverhaltens von Testpersonen im Musikladen zeigt eine deutliche Tendenz der Probanden zum „shopping around“ im Gegensatz zum „shopping for“. Kunden durchstöbern ganze Regale und lassen sich vom Angebot inspirieren. Dabei ist der Explorationsprozess nicht vollständig unstrukturiert: So beschränken sich viele Probanden auf CDs, die vom Laden-Besitzer einem den Kunden interessierenden Genre zugeordnet wurden.

Laplane und Downie unterstützen Cunninghams Beobachtungen und stellen in einer ähnlichen Studie die Bedeutung des explorativen Zugangs zur Erschließung einer Musikkollektion heraus [LD06].

1.2.2 Explorations-Unterstützung aktueller Musikportale

Aktuelle Musik-Internet-Portale gestatten zumeist nur einen sehr eingeschränkten explorativen Zugang zu ihrem Musikkatalog. Die folgende Zusammenstellung listet einige der Zugriffshilfen auf, die ein ungerichtetes Stöbern im Angebot erleichtern sollen. Die meisten dieser Verfahren zielen auf eine schnelle Erreichbarkeit oder Gruppierung zueinander *ähnlicher* Stücke ab.

- *Item-to-Item-Empfehlungen*

Die Beobachtung des Kauf- und Navigationsverhaltens hinreichend vieler anderer Online-Portal-Nutzer ermöglicht die Etablierung von kollaborativ generierten Verweisstrukturen zwischen Künstlern, Alben oder Titeln. Ausgehend von einem bekannten Album kann der Kunde so zu weiteren Produkten gelangen, die von anderen Benutzern in ähnlichem Kontext besucht oder gekauft wurden und somit als ähnlich angenommen werden.

- *Genre-Listen*

Alben und/oder Künstler werden von Musik-Portal-Betreibern gemäß einer mehr oder weniger komplexen Genre-Taxonomie klassifiziert. Amazon.com erlaubt die Zuweisung mehrerer Genres zu einem Album, Last.fm verzichtet ganz auf vorgegebene Genre-Bezeichner und setzt stattdessen auf die Entwicklung einer Folksonomy aus frei assoziierbaren Tags.

Nutzer können diese Annotation(en) als Filter-Kriterium auf den Musikkatalog verwenden und so z.B. alle angebotenen Alben des Genres „Indie Rock“ betrachten.

- *Benutzerlisten*

Amazon.com erlaubt es seinen Benutzern, Alben in persönlichen Listen zu gruppieren. Diese Listen können eine Bezeichnung tragen und für andere Benutzer zugreifbar sein. Ferner wird von der Darstellungsseite eines Albums auf die Listen verwiesen, die das Album enthalten.

Unter der Annahme, dass Benutzer vornehmlich ähnliche Musik zu Listen gruppieren, erlaubt dies dem Kunden eine Navigation zu verwandten Produkten.

- *Neuigkeiten und Charts*

Die meisten Online-Portale stellen Neuerscheinungen und populäre Produkte auf eigenen Übersichtsseiten zusammen und bieten dem Kunden somit einen initialen Einstiegspunkt in einen Explorationsprozess.

1.2.3 Kritik aktueller Explorations-Unterstützung

Damit der Benutzer in der Weite des Musikkataloges nicht die Orientierung verliert und sein Durchstöbern nicht zu einem zufälligen Zugriff degradiert, muss eine Explorationsumgebung versuchen, Interesse und Ordnungsprinzip des Kunden zu antizipieren und die Musikkollektion gemäß seinem Ähnlichkeitsempfinden strukturiert aufbereiten.

Oben beschriebene Zugriffshilfen ermöglichen zwar eine Navigation entlang verschiedener Ähnlichkeitsbegriffe, deren Integration, die eine einheitliche Strukturierung der Musiksammlung ermöglichen würde, bleibt jedoch aus.

Ferner gelangt der Benutzer mit jedem Schritt in einen von der bisherigen Navigationshistorie unabhängigen Kontext mit Verweisen zu weiteren Produkten (die möglicherweise bereits besucht wurden). Bisherige Explorationsunterstützung beschränkt sich somit auf das Angebot einer *einschrittigen* Navigationshilfe. Dies erschwert die Erkundung umfangreicherer Musiksammlungen, da sie vom Benutzer eine klare Explorationsstrategie und ein Erinnern der getroffenen Navigationsentscheidungen verlangt.

Aktuelle Musik-Portale verzögern die akustische Wahrnehmung der Musik bis zum Ende der Navigationskette, wenn sie überhaupt ein Vorhören des Materials erlauben. So muss der Benutzer z.B. im Falle einer Exploration entlang der Genre-Kategorisierung die Bezeichner des Anbieters interpretieren und für sich in eine musikalische Stilvorstellung umsetzen können. Dieser unnatürliche Bruch zwischen textueller Beschreibung und akustischem Medium erschwert das Treffen einer informierten Navigationsentscheidung durch den Kunden deutlich. Wünschenswert ist stattdessen ein „Klangraum“, der die Musik selbst darstellt und in dem sich der Benutzer akustisch orientieren und bewegen kann.

1.2.4 Nutzer-adaptive Exploration

Wie beschrieben ist die Verfügbarkeit eines Ähnlichkeitsbegriffes, der die Strukturierung einer Musikkollektion ermöglicht, zentral für die Qualität einer Explorationsumgebung. Wie Cunningham in einer Untersuchung der Organisationsprinzipien persönlicher Musiksammlungen vornehmlich bei Jugendlichen feststellt, ist dieser Ähnlichkeitsbegriff jedoch subjektiv sehr verschieden [CJJ04]. Manche Nutzer strukturieren ihre Musik nach Erscheinungs- oder Kaufdatum, andere gruppieren ihre CDs gemäß persönlicher Genre-Cluster. Die Vielfältigkeit von Genre-Definitionen ist in einer Reihe von Publikationen insbesondere auch im Hinblick auf die Evaluierbarkeit von Genre-Klassifikatoren beschrieben worden [PC00, SZM06, AP03, SCBG08].

Des weiteren ist die Bedeutung verschiedener Musikmerkmale abhängig von der betrachteten Musikkollektion: So sind Tempo und Beat-Stärke innerhalb einer DJ-Sammlung wichtige Stückcharakteristika, in einer Klassik-Kollektion aber nahezu unbedeutend.

Diese Beobachtungen legen die Forderung nach einer flexiblen Explorationsumgebung nahe, die eine anpassbare Gewichtung unterschiedlicher Musikmerkmale zur Strukturierung der Sammlung gestattet.

Die Kombination unterschiedlicher Ähnlichkeitsmaße wird in einer Reihe von Publikationen untersucht [SLT07, Jac06, LST04, BPV04, PDW03a]. Baumann erlaubt dem Benutzer durch Bewegung eines virtuellen Joysticks die Festlegung einer Linearkombination aus inhaltsbasierter Ähnlichkeit, Stil-Ähnlichkeit (die aus textueller Ähnlichkeit charakterisierender musikalischer Begriffe beschreibender Webseiten abgeleitet wird) und textueller Ähnlichkeit der Liedtexte [BPV04]. Pampalk gestattet eine interpolierte Überblendung zwischen abgeglichenen selbstorgansierenden Karten, die die betrachtete Stückmenge gemäß unterschiedlicher Ähnlichkeitsbegriffe anordnen [PDW03a].

In beiden Fällen muss der Benutzer selbständig die Gewichtung der einzelnen Ähnlichkeitsmaße vornehmen. Mit steigender Anzahl verfügbarer Maße ist dies ein komplexer Prozess, der vom Benutzer ein Verständnis der Bedeutung der angebotenen Ähnlichkeitsbegriffe verlangt. In dieser Arbeit wird daher der alternative Ansatz eines adaptiven Systems verfolgt, das aus Restrukturierungen der Explorationslandschaft, die der Benutzer vornehmen kann, eine Neugewichtung der Ähnlichkeitsmaße lernt und sich somit schrittweise der Strukturierungsstrategie des Benutzers anpasst.

1.2.5 Zielsetzung

Ziel dieser Arbeit ist die Entwicklung und Evaluierung einer Explorationsumgebung für Musikkollektionen, die eine intuitive Orientierung und Navigation in der Musiksammlung gestattet.

Hierbei sollen insbesondere die folgenden Forschungskernfragen untersucht werden:

- *Akustische Orientierung*

Wie lässt sich der Medienbruch zwischen Stücken und deren textueller Beschreibung vermeiden?

Wie können optische Strukturierung und akustische Eindrücke zu einer intuitiven multi-modalen Explorationsumgebung kombiniert werden?

Inwieweit hilft eine akustische Umgebung bei Orientierung und Navigation in der Musikkollektion?

- *Benutzer-adaptives Ähnlichkeitsmaß*

Welche Einzelähnlichkeitsmaße sind relevant?

Welche initiale Kombination von Ähnlichkeitsmaßen ist sinnvoll?

Wie kann aus Restrukturierungen des Explorationsumgebung durch den Benutzer auf eine Veränderung der Gewichtung von Ähnlichkeitsmaßen geschlossen werden?

Aufgrund der oben beschriebenen wachsenden Größe persönlicher Musiksammlungen, stellen diese – neben dem Anwendungsfeld von Online-Musik-Portalen – eine ebenso bedeutende Domäne für die zu entwickelnde Explorationsumgebung dar.

Eine intuitive Umgebung, die die Stücke einer Kollektion in einer an den Benutzer angepassten Weise strukturiert, ist dabei mehr als Explorationsumgebung, sondern kann auch Informationsbedürfnisse der gezielten Suche erfüllen, wenn die Suchparameter zwar fest, aber nicht textuell artikulierbar sind.

1.3 Repräsentationsformen von Musik

Zur Abgrenzung der Aufgabenstellung und Einordnung der Arbeit in das Forschungsumfeld des Music Information Retrieval (MIR) werden im folgenden drei Repräsentationsklassen beschrieben, die Musik auf unterschiedlichen Abstraktionsniveaus darstellen. Ferner werden Arbeiten skizziert, die einen computergestützten Wechsel des Abstraktionsniveaus ermöglichen.

1.3.1 Musik als Noten

Bis an den Anfang des 20. Jahrhunderts bestand die Musikindustrie neben den Kunstschaufenden selbst primär aus Verlegern, die Musik in Form von Notenblättern publizierten, der einzigen damaligen Möglichkeit, Töne und Klänge festzuhalten. Abbildung 1.3 zeigt mit den ersten beiden Takten der 13. Invention von Johann Sebastian Bach ein Beispiel für diese klassische westliche Musik-Notation.

Diese *symbolische* Repräsentation von Musik abstrahiert nicht nur von der konkreten Interpretation des Stückes durch einen Musiker, die sich z.B. in Agogik und Dynamik äußert, sondern auch vom konkreten Klang der Instrumente, deren Positionierung im Raum, den akustischen Gegebenheiten usw. Sie beschränkt sich vielmehr im Wesentlichen auf die Festlegung von Tonhöhen und -längen. Zusätzliche Spielanweisungen legen Tempo(-veränderungen), Dynamik, Artikulation und Instrumentierung fest, bleiben aber mitunter recht vage und erlauben einen gewissen Interpretationsspielraum.



Abbildung 1.3: Die ersten beiden Takte der 13. Invention von Johann Sebastian Bach als Notentext

Daher ist symbolische repräsentierte Musik vor allem zur Beschreibung klassischer Musik geeignet; moderne Popproduktionen ziehen ihren Reiz zum Teil aus charismatischen Stimmen, Soundeffekten und Geräuschen, die in Form von Noten nicht oder nur sehr unspezifisch darstellbar sind.

Eine erste Sprache zur Speicherung symbolischer Musik im Rechner war *Humdrum* [Hur97]. Humdrum spezifiziert ein eindeutiges Benennungsschema für Tonhöhen, Pausen und Tonlängen, die im einfachsten Anwendungsfall zu einer monophonen Linie sequentiell aneinandergereiht werden. Die Möglichkeit, in mehreren Spalten unabhängige „Tonströme“ parallel anzugeben, erlaubt die Repräsentation polyphoner Musik.

Die Beliebtheit von Humdrum in ersten Arbeiten des Music Information Retrieval (z.B. [vSDR01, Kor01]) lässt sich vor allem durch die Verfügbarkeit einer ausgereiften Toolbox zur Bearbeitung und Analyse von Humdrum-Daten zu erklären [Hum], die Anfragen im Wesentlichen auf reguläre Ausdrücke zurückführt und diese mittels des Unix-Tools *grep* auswertet.

Heute gilt Humdrum als veraltet. Noten werden meist in proprietären Formaten der dominierenden Notensatzprogramme wie Finale¹, Sibelius² (oder in Deutschland auch capella³) repräsentiert.

Als gemeinsames Austauschformat hat sich MusicXML [Goo01] etabliert. Der MusicXML-Standard erlaubt eine fast vollständige Kodierung aller Elemente notierter Musik (inkl. Instrumentierung und Spielanweisungen) und ermöglicht damit nicht nur eine weitestgehend verlustfreie Konversion zwischen obigen Notationsprogrammen, sondern bietet auch die Basis für eine Ausführung komplexer MIR-Anwendungen. Abbildung 1.4 stellt Ausschnitte aus der MusicXML-Datei zur Repräsentation des Notentextes aus Abbildung 1.3 dar.

Aufgrund des Abstraktionsniveaus symbolischer Musik sind die typischen Fragestellungen solcher Anwendungen musikwissenschaftlich motiviert. So geht es beispielsweise um die Analyse der Verarbeitung eines Motives im Verlauf eines Werkes [Wey02] oder um die Identifikation von Melodie-Gemeinsamkeiten zum Zwecke der Plagiarismus-Erkennung [HS72]. Beide Anwendungsfelder basieren auf einer Ähnlichkeitsfunktion für symbolische Musik, die auf Teile innerhalb eines Stückes oder zwischen Werken angewendet wird. Hier haben sich vor allem drei Ansätze etabliert:

¹<http://www.finalemusic.com>

²<http://www.sibelius.com>

³<http://www.whc.de/capella.cfm>

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE score-partwise PUBLIC "-//Recordare//DTD MusicXML 1.1 Partwise//EN"
"http://www.musicxml.org/dtds/partwise.dtd">
<score-partwise version="1.1">
  <movement-title>Invention Nr. 13</movement-title>
  ...
  <part-list>
    <score-part id="P1">
      <part-name>Piano</part-name>
      <score-instrument id="P1-I1">
        <instrument-name>Piano</instrument-name>
      </score-instrument>
    </score-part>
  </part-list>
  <!--=====-->
  <part id="P1">
    <measure number="1">
      <attributes>
        <divisions>4</divisions>
        <key>
          <fifths>0</fifths>
          <mode>minor</mode>
        </key>
        <time symbol="common">
          <beats>4</beats>
          <beat-type>4</beat-type>
        </time>
        <staves>2</staves>
      </attributes>

      <note>
        <rest/>
        <duration>1</duration>
        <voice>1</voice>
        <type>16th</type>
        <staff>1</staff>
      </note>
      ...
      <note>
        <pitch>
          <step>G</step>
          <alter>1</alter>
          <octave>4</octave>
        </pitch>
        <duration>2</duration>
        <voice>1</voice>
        <type>eighth</type>
        <accidental>sharp</accidental>
        <stem>down</stem>
        <staff>1</staff>
        <beam number="1">continue</beam>
        <notations>
          <articulations>
            <staccato placement="above"/>
          </articulations>
        </notations>
      </note>
      ...
    </measure>
  </part>
  <!--=====-->
</score-partwise>

```

Abbildung 1.4: Ausschnitt aus Music-XML-Repräsentation der 13. Invention

1. *Reduktion auf Fuzzy-String-Matching*

Monophone Tonfolgen werden als N-Gramme in einer Datenbank gespeichert. Um die meist gewünschte Transpositions- und Augmentations-/Diminutions-Invarianz zu erreichen, kodiert man in einem solchen N-Gramm lediglich die Intervalle zwischen benachbarten Tönen sowie deren Tonlängenverhältnisse. Durch diese Reduktion auf String-Fragmente lassen sich mit Hilfe bekannter Matching-Algorithmen zu einer Anfrage ähnliche Abschnitte in der gespeicherten Tonfolge finden.

Dieses Verfahren ist sehr effizient, muss jedoch zur Unterstützung der Suche in polyphonem Material alle möglichen monophonen Linien extrahieren, was zu einer mit dem Grad der Polyphonie exponentiell steigenden Datenbankgröße und zur Tendenz der Rückgabe falsch-positiver Suchergebnisse führt [DR03].

2. *Earth-Movers-Distance*

Ein Ansatz, der eine elegante Modellierung musikwissenschaftlichen Wissens über Melodie-Ähnlichkeit ermöglicht, basiert auf der Earth-Movers-Distance. Dabei wird der Abstand zwischen zwei monophonen Linien als minimale Summe der Kosten zur Überführung der Töne der einen Melodie in die Töne der anderen Melodie definiert. Typke verwendet als Kostenmaß für die Transformation eines Tones eine Kombination aus Längen-, Anfangszeitpunkts- und Höhendifferenz [TGV⁺03, TWV04].

3. *Stochastische Harmonie-Modellierung*

Obige Ansätze sind primär für den Vergleich (quasi-)monophoner Musik geeignet. Pickets entwickelt in [Pic04] ein Ähnlichkeitsmaß für polyphone Musik, das nicht auf Einzelton-Verläufen basiert, sondern Stücke als stochastische Modelle ihrer Harmonik auffasst.

Hierzu wird ein gegebenes polyphones Stück auf eine Sequenz von Mengen gleichzeitig erklingender Töne reduziert. Geht man von einer chromatischen Tonskala aus und abstrahiert von der Oktavlage, existieren $2^{12} = 2048$ verschiedene solcher „Simultanities“. Eine Simultaneity-Folge wird als Beobachtungssequenz eines Hidden-Markov-Modells verstanden, dessen verborgene Zustände den $2 \cdot 12$ Dur- bzw. Mollldreiklängen entsprechen. Als initiale Übergangswahrscheinlichkeiten zwischen diesen Zuständen wird ihr Abstand im Quintenzirkel angenommen. Mit Hilfe des Baum-Welch-Algorithmus werden diese Übergangswahrscheinlichkeiten zusammen mit den Beobachtungswahrscheinlichkeiten jedes verborgenen Zustandes aus den Simultanities eines Stückes erlernt.

Als Distanz zwischen einem so modellierten Stück und einer Anfrage wird schließlich die Wahrscheinlichkeit herangezogen, mit der das gegebene Hidden-Markov-Modell die Simultaneity-Sequenz der Anfrage erklärt.

Das größte Problem dieses Ansatzes ist seine Beschränkung auf ein Hidden-Markov-Modell erster Ordnung, in dem lediglich Übergangswahrscheinlichkeiten von einer Harmonie zur nächsten erfasst werden. Hierdurch abstrahiert die Repräsentation nach Pickets weitestgehend vom zeitlichen Ablauf der Harmonik und kann z.B. nicht zwischen den Akkordsequenzen (C F G Am G D G C) und (C F G D G Am G C) unterscheiden.

1.3.2 Musik als Steuersignal

Musik lebt von der Interpretation durch den Musiker, die sich in feinen bewussten und unbewussten Abweichungen von einer mathematisch exakten Ausführung notierter Töne und Spielanweisungen äußert. Die im letzten Abschnitt vorgestellte Repräsentationsform abstrahiert bewusst von diesen Ungenauigkeiten und fokussiert (auch aus Gründen der besseren Lesbarkeit) auf den interpretationsunabhängigen „Kern“ der Musik.

Ist man an einem menschlich-authentischeren Musikgenuss interessiert oder will man z.B. die einen Musiker charakterisierenden Interpretationseigenarten musikwissenschaftlich identifizieren und analysieren, ist diese Reduktion kontraproduktiv. Anstatt dessen bietet sich die Darstellung in einer Repräsentation an, die sich aus dem Übertragungsprotokoll zwischen elektronischen Klangerzeugern, dem bereits 1983 spezifizierten *MIDI-Protokoll*, ableitet und daher *MIDI-Datei-Format* genannt wird.

Der MIDI-Standard legt eine Reihe von Steuerkommandos fest, die von einem MIDI-Produzenten (wie einem Keyboard oder einem MIDI-Datei-Spieler) generiert und einem MIDI-Empfänger (z.B. einem tonerzeugenden Synthesizer) konsumiert werden. Die wichtigsten dieser Steuerkommandos sind die Note-On- sowie die Note-Off-Meldung, die zusammen mit der als Velocity bezeichneten Anschlagsintensität Betätigung und Loslassen einer Taste auf einem Keyboard kodieren. Darüber hinaus lassen sich eine Reihe weiterer Kommandos übermitteln, die die Tonerzeugung steuern, z.B. die Festlegung eines Klangprogramms, Lautstärke, Panoramaposition oder (in begrenztem Maße) Effekteinstellungen.

MIDI ist ein sequenzielles Protokoll. Zur Adressierung von Steuerbefehlen an verschiedene unabhängig spielende Instrumente ist jedem Kommando einer von 16 möglichen Kanälen zugeordnet. Multi-timbrale Synthesizer können anhand dieser Kanalzahl eintreffende Steuerbefehle an separate Klangerzeugungsmodule weiterleiten.

Trotz seines klaren Ursprungs in der Kodierung des Spiels auf Tasteninstrumenten gibt es inzwischen eine Reihe anderer spielbarer MIDI-Produzenten, z.B. elektronische Schlagzeuge, MIDI-Gitarren oder MIDI-Blaswandler, die die vielfältigen Artikulationsmöglichkeiten und Zwischen-Tonhöhen von Blasinstrumenten durch spezielle Steuerkommandos repräsentieren.

Das MIDI-Datei-Format erweitert die Protokollspezifikation im Wesentlichen um die Möglichkeit zur zeitlichen Ordnung und Fixierung von MIDI-Meldungen. Ferner erlaubt es die Integration einer Reihe von Metadaten, die für die Wiedergabe irrelevant sind.

Somit ermöglicht eine MIDI-Datei die Beschreibung der Interpretation eines notierten Stückes. Durch die Konzentration auf die Ausführung durch einen Musiker findet jedoch auch eine Informationsreduktion statt: So lassen sich Tonart und Taktart nicht mehr direkt aus den Note-On- und Note-Off-Meldungen ablesen; enharmonische Verwechslungen (z.B. zwischen F# und Ges) werden in MIDI durch die gleiche Tastennummer kodiert und sind damit (im Gegensatz zum Notendruck) nicht mehr unterscheidbar; das polyphone Gefüge von unabhängigen Einzelstimmen auf mehrstimmig spielbaren Instrumenten wie einem Klavier ist nicht beschreibbar.

MIDI legt zwar die Interpretation notierter Musik fest, bleibt bei der Spezifikation des resultierenden Klanges jedoch vage und beschränkt sich auf die textuelle Benennung des Instrumentes. Je nach genutztem MIDI-Wiedergabe-Gerät klingen MIDI-Datei daher zum Teil recht unterschiedlich.

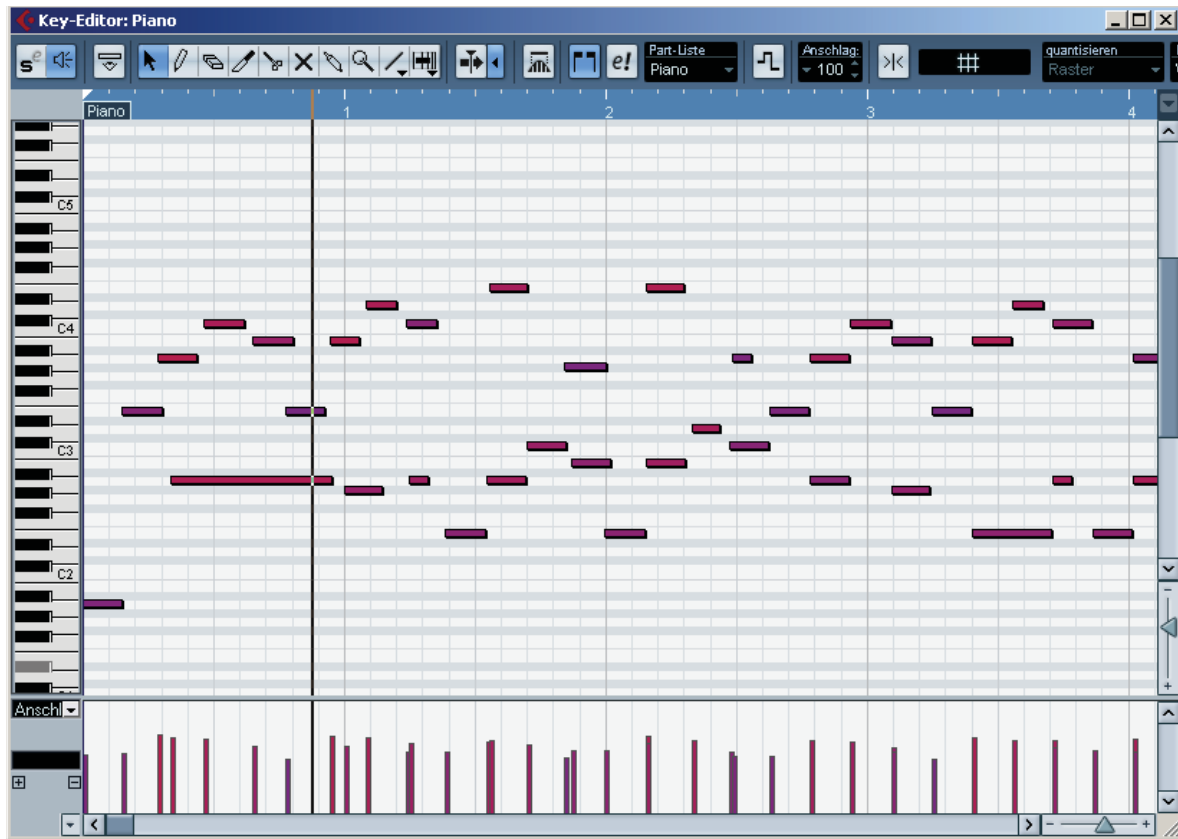


Abbildung 1.5: Interpretation der 13. Invention als MIDI-Datei in Piano-Roll-Darstellung. Screenshot aus der Musikproduktionsumgebung Cubase.

Diese Abstraktion ist Grund für die Kompaktheit von MIDI-Dateien und die Klassifikation von MIDI als symbolisches Repräsentationsformat. Vor der Etablierung von Kompressions-Formaten, die einen ressourcenschonenden Umgang mit Klangdateien ermöglichen, waren MIDI-Dateien das primäre Format zum Austausch von Musik im Internet. Eine Renaissance hat das Format zudem zeitweise als Beschreibungssprache polyphoner Handy-Klingeltöne erfahren.

Abbildung 1.5 zeigt MIDI-Daten einer Interpretation der 13. Invention in einer sog. Piano-Roll-Darstellung, bei der die horizontale Achse den zeitlichen Verlauf kennzeichnet und die vertikale Achse die Höhe von Tönen spezifiziert, die als Balken im Diagramm symbolisiert werden. Im unteren Teil des Fensters wird die Anschlagsstärke der Töne dargestellt. Man erkennt deutlich unterschiedliche Tonlängen, die sich nicht allein aus dem Notentext ergeben, sondern durch die Interpretation des Musikers zu erklären sind. Abbildung 1.6 stellt einen Ausschnitt der gleichen Daten in Form einer sequenziellen MIDI-Liste dar.

1.3.3 Musik als Wellenform

Die für den Nicht-Musiker bei weitem bedeutendste Repräsentationsform fasst Musik als zumeist zweikanaliges digitalisiertes Signal von Luftdruckschwankungen auf. Aufgrund der

L	Typ	Anfang	Ende	Länge	Wert 1	Wert 2	Wert 3	Kanal
	♪ Note	1. 1. 1. 0	1. 1. 1.118 0. 0. 0.118	A1	66	64	1	
	♪ Note	1. 1. 1.117	1. 1. 2.117 0. 0. 1. 0	E3	68	64	1	
	♪ Note	1. 1. 2.101	1. 1. 3.101 0. 0. 1. 0	A3	90	64	1	
	♪ Note	1. 1. 3. 22	1. 2. 3. 22 0. 1. 0. 0	A2	87	64	1	
	♪ Note	1. 1. 4. 0	1. 2. 1. 0 0. 0. 1. 0	C4	85	64	1	
	♪ Note	1. 2. 1. 25	1. 2. 2. 25 0. 0. 1. 0	B3	77	64	1	
	♪ Note	1. 2. 2. 0	1. 2. 3. 0 0. 0. 1. 0	E3	61	64	1	
	♪ Note	1. 2. 3. 16	1. 2. 3.104 0. 0. 0. 88	B3	89	64	1	
	♪ Note	1. 2. 3. 57	1. 2. 4. 53 0. 0. 0.116	G#2	76	64	1	
	♪ Note	1. 2. 4. 0	1. 2. 4. 96 0. 0. 0. 96	D4	86	64	1	
	♪ Note	1. 3. 1. 0	1. 3. 1. 93 0. 0. 0. 93	C4	69	64	1	
	♪ Note	1. 3. 1. 9	1. 3. 1. 68 0. 0. 0. 59	A2	80	64	1	
	♪ Note	1. 3. 1.117	1. 3. 2.117 0. 0. 1. 0	E2	70	64	1	
	♪ Note	1. 3. 3. 0	1. 3. 4. 0 0. 0. 1. 0	A2	82	64	1	
	♪ Note	1. 3. 3. 6	1. 3. 4. 4 0. 0. 0.118	E4	83	64	1	
	♪ Note	1. 3. 3.117	1. 3. 4.117 0. 0. 1. 0	C3	78	64	1	
	♪ Note	1. 3. 4.108	1. 4. 1.118 0. 0. 1. 10	G#3	63	64	1	

Abbildung 1.6: Interpretation der 13. Invention als MIDI-Datei in Listen-Darstellung. Screenshot aus der Musikproduktionsumgebung Cubase.

mittels stereophoner Wiedergabe näherungsweise Reproduzierbarkeit des aufgenommenen Schallfeldes beim Hörer handelt es sich hierbei um das detaillierteste der vorgestellten Abstraktionsniveaus. Abbildung 1.7 zeigt den bekannten Beginn der 13. Invention als Wellenform einer Interpretation auf dem Klavier.

Die Repräsentation von Musik als Klang erlaubt zwar eine authentische Wiedergabe, erschwert gleichzeitig jedoch die Extraktion musikalischer (und vor allem musikwissenschaftlich interessanter) Parameter. Abschnitt 1.3.4 gibt einen kurzen Überblick über Verfahren, die in einer Art „Reverse-Engineering“-Prozess versuchen, gespielte Einzeltöne im Gesamtklang zu identifizieren und für eine weitere Analyse verfügbar zu machen.

Wie Aucouturier [AF04] argumentiert, ist das Timbre eines Stückes für Nicht-Musiker ein dominantes Merkmal für einen perzeptuellen Ähnlichkeitsbegriff. Es scheint plausibel, dass Kunden eines Online-Musik-Portales primär nach „ähnlich klingender“ Musik suchen. Diese Ähnlichkeit erfordert die Repräsentation von Musik als Wellenform.

Aus der Detailliertheit des Klangsignals folgt ein erhöhter Speicherbedarf. Legt man die Spezifikation einer Audio-CD (Abtastrate 44.1 kHz, 16-Bit-Quantisierung, 2 Kanäle (Stereo)) zugrunde, beansprucht eine Sekunde Musik 172 KByte. Ein typischer drei Minuten langer Pop-Song benötigt damit mehr als 30 MByte. Dies macht klar, dass erst die Verfügbarkeit von akustisch transparenten Kompressionsverfahren wie MP3 eine effiziente Verteilung von Musik als Klangdatei ermöglichen.

Die aktuellere MIR-Literatur konzentriert sich fast ausschließlich auf die Analyse von Klangsignalen, wandelt hierzu die Wellenform von Luftdruckschwankungen jedoch häufig in eine spektrale Darstellung um, die eine einfachere Identifikation dominanter Frequenzen (und damit (Partial-)Töne) in einem betrachteten Zeitfenster gestattet. In den meisten Fällen wird solch eine spektrale Darstellung durch die Diskrete Fourier-Transformation berechnet (für eine genauere Einführung vergleiche z.B. [OSB99]):

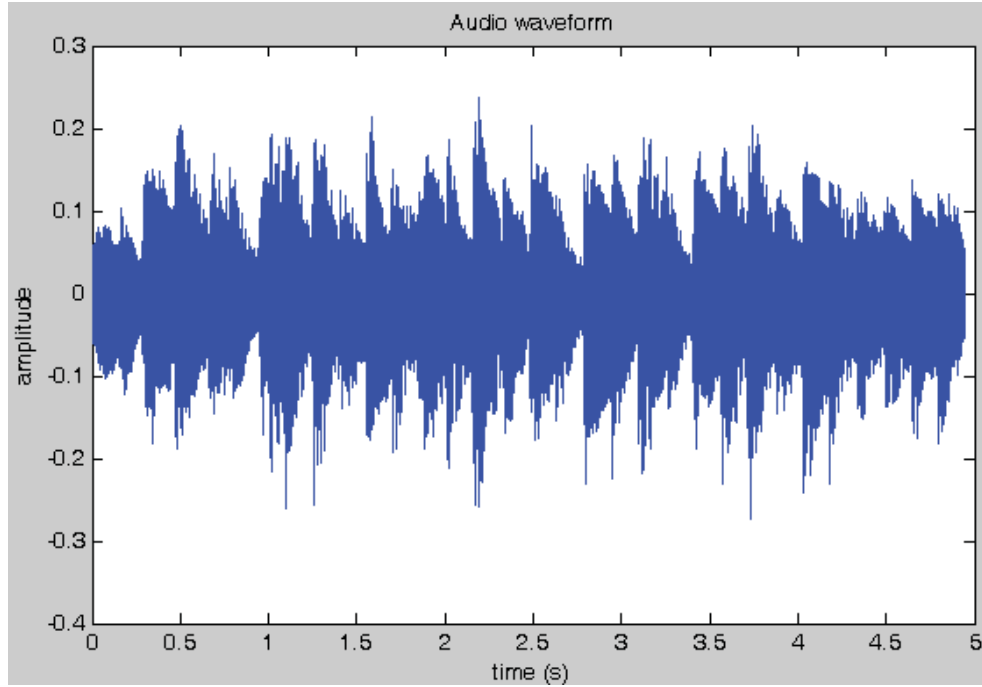


Abbildung 1.7: Darstellung einer Interpretation der 13. Invention auf dem Klavier als Wellenform.

Ausgangspunkt ist ein in N äquidistanten Punkten mit der Samplingfrequenz f_S abgetastetes Signal $\vec{x} = (x[0], \dots, x[N-1])^T$. Für die folgende Darstellung verzichten wir auf eine Berücksichtigung der Quantisierung der Signalwerte und gehen von $x[i] \in \mathbb{R}$ bzw. $x[i] \in \mathbb{C}$ aus.

Damit lässt sich $\vec{x}$ als N -dimensionaler Vektor in $\mathbb{C}^N$ interpretieren.

Wie sich leicht überprüfen lässt, stehen die Vektoren

$$\vec{b}_k := \frac{1}{\sqrt{N}}(\Omega_N^{0k}, \Omega_N^{1k}, \dots, \Omega_N^{(N-1)k})^T \text{ für } k = 0..N-1$$

mit

$$\Omega_N = e^{\frac{2\pi j}{N}} \text{ (erste } N\text{-te Einheitswurzel)}$$

paarweise orthogonal zueinander und haben die Länge 1. Damit handelt es sich um eine alternative Basis des $\mathbb{C}^N$, in der sich das Signal $\vec{x}$ als

$$\vec{x} = \sum_{k=0}^{N-1} X[k] \cdot \vec{b}_k$$

umformulieren lässt. Die komplexwertigen $X[k]$ dieser Basis werden als *Fourier-Koeffizienten* bezeichnet. Ihr Betrag gibt die Intensität der Frequenz $\frac{k}{N} \cdot f_S$ im Signal an.

Aufgrund der Orthonormalität von $\{\vec{b}_0, \dots, \vec{b}_{N-1}\}$ lassen sich diese $X[k]$ einfach als Skalarprodukt berechnen:

$$X[k] = \langle \vec{x}, \vec{b}_k \rangle = \sum_{i=0}^{N-1} x[i] \cdot \overline{b_k[i]}.$$

In der Literatur verzichtet man meist auf den Faktor $\frac{1}{\sqrt{N}}$ in b_k und damit auf deren Normalität. Anstatt dessen normalisiert man die $X[k]$ mit dem Faktor $\frac{2}{N}$ und erhält somit über verschiedene Werte von N vergleichbare Intensitäten.

Unter der Voraussetzung eines reellwertigen Signals $\vec{x}$ gilt $X[\frac{N}{2} - k] = X[\frac{N}{2} + k]$ für gerade N , d.h. die Koeffizientenfolge ist symmetrisch und nur die erste Hälfte enthält unabhängige Informationen. Dies steht in Einklang mit dem Nyquist-Shannon'schen Abtast-Theorem, das eine Samplingfrequenz von mindestens f_S für ein Signal mit einer Grenzfrequenz von $\frac{1}{2}f_S$ fordert.

Zur effizienten Berechnung der Diskreten Fourier-Transformation wird in der Regel der Algorithmus von Cooley and Tukey [CT65] verwendet, der für Zweierpotenzen von N den quadratischen Aufwand auf $O(N \log N)$ reduziert.

Zumeist ist man nicht an dem Frequenzspektrum über das gesamte Stück interessiert, sondern möchte die Intensität bestimmter Töne im zeitlichen Verlauf des Signals bestimmen. Hierzu schneidet man die Samplefolge in (möglicherweise überlappende) Fenster und führt das Teilsignal jedes Fensters einer Spektralanalyse zu. Zur Vermeidung von Rand-Effekten multipliziert man die betrachtete Samplefolge zuvor mit einer Fenster-Funktion (im MIR-Kontext wird meist eine Hanning- oder Hamming-Funktion verwendet). Die Frequenzauflösung der Fourier-Transformation hängt mit $\frac{1}{N}$ direkt von der Anzahl der Samples ab. Daher muss ein Kompromiss zwischen Lokalität und Auflösung einer Fourier-Darstellung gefunden werden.

Abbildung 1.8 zeigt eine Spektralanalyse des Signals aus Abbildung 1.7. Horizontale Zeit- und vertikale Frequenz-Achse spannen eine Ebene auf, in der die Intensität einer Frequenz zu einem bestimmten Zeitpunkt im Signal durch die Färbung der entsprechenden Pixelposition dargestellt wird.

Aufgrund der Bedeutung in Online-Portalen und der Fokussierung auf Explorationsumgebungen, die auch von Nicht-Musikern bedienbar sind, beschränkt sich diese Arbeit auf die Betrachtung von Musik als akustisches Signal.

1.3.4 Wechsel des Abstraktionsniveaus

Im folgenden werden überblicksartig einige Ansätze zur Konversion einer Musikrepräsentation in ein benachbartes Abstraktionsniveau diskutiert.

Noten und Interpretation

Die Umwandlung einfacher notierter Musik in eine Sequenz aus Note-On- und Note-Off-Meldungen ist trivial. Die Aufgabe wird komplexer, wenn auch Spielanweisungen wie Dynamik, Phrasierung, Artikulation, Tempoänderungen etc. berücksichtigt werden müssen. Dennoch leisten fast alle kommerziell bedeutsamen Notensatzprogramme diese Konvertierung problemlos.

Eine so erzeugte Steuersignalsequenz klingt jedoch unnatürlich mechanisch und leblos. Insbesondere die Interpretation klassischer Musik ist geprägt von der Anwendung einer Reihe impliziter Darbietungskonventionen, die der Musiker durch seine Ausbildung erlernt hat und deren Berücksichtigung Komponist und Hörer erwarten.

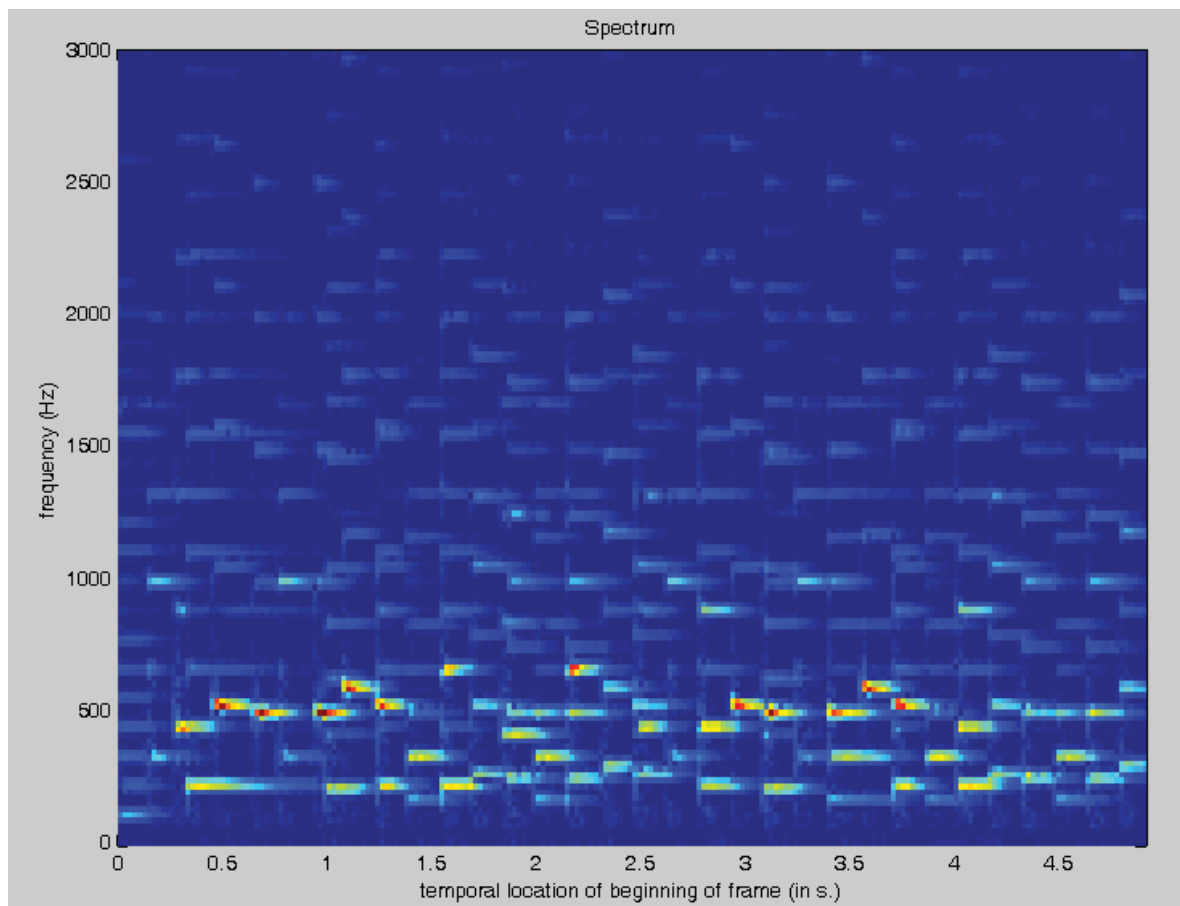


Abbildung 1.8: Darstellung einer Interpretation der 13. Invention auf dem Klavier als Spektrogramm. Man erkennt deutlich Grundfrequenzen und Partialtöne der angeschlagenen Klavier-Saiten.

Die Modellierung eines prädiktiven Gestaltungs-Modells ist daher Thema aktiver Forschung [ALdM01, LdMA02, JFB02, Hon03]. Aktuelle Ergebnisse werden ferner in regelmäßigen Abständen im Rahmen des RENCON⁴ (Performance Rendering Contest for Computer Systems) gegenübergestellt.

Im folgenden werden beispielhaft zwei Ansätze zur Modellierung von impliziten Interpretationskonventionen skizziert.

Widmer et al. unternehmen in einem viel zitierten Projekt den Versuch, diese Gesetzmäßigkeiten mit Hilfe Induktiver Logischer Programmierung datengetrieben zu gewinnen [WDG⁺03]. Grundlage dieses Data Mining-Prozesses ist die Einspielung von dreizehn Klaviersonaten von Mozart durch einen Konzertpianisten auf einem speziell präparierten Flügel, der die Generierung von MIDI-Daten aus Tastenschlägen erlaubt. Der so erzeugte Steuersignal-Strom wird anschließend mit dem vorliegenden Notentext verglichen. Widmer et al. induzieren Zusammenhänge zwischen Interpretationsregeln auf zwei Ebenen: Zum einen setzen sie eine Reihe von Attributen, die intrinsische Eigenschaften (wie z.B. Dauer, metrische Position, etc.) und den lokalen Kontext *einer Note* charakterisieren, zu Gestaltungsparametern wie Dynamik, Tempoveränderung und Artikulation in Bezug. Zum anderen leiten sie Regeln zur Interpretation größerer Phrasierungseinheiten ab. Diese induzierten Regeln identifizieren lediglich Situationen, in denen eine Abweichung vom exakten Notentext vorzunehmen ist. Die Intensität einer solchen Abweichung wird durch eine k-Nearest-Neighbour-Suche unter allen Anwendungsfällen der Regel in der Trainingsmenge quantifiziert.

Damit gewinnen Sie ein einfaches Modell, das nach [WDG⁺03] erstaunlich gut die Gestaltung ungesehener Klavier-Notentexte vorhersagt und somit authentisch wiedergeben kann.

Hazan et al. beschäftigen sich mit der Interpretation von Jazz-Standards auf dem Saxophon [HGR06]. Im Gegensatz zum obigen Anwendungsfeld sind die gestalterischen Freiheiten des Interpreten hier deutlich umfangreicher und umfassen neben den bereits beschriebenen u.a. auch die Einfügung kleiner Zwischentöne oder die Ornamentierung langer Töne. Zur Modellierung dieser Abweichungen werden sog. Regressionsbäume verwendet, die abhängig vom musikalischen Kontext einer Note eine interpretatorische Reaktion vorhersagen. Hazan et al. gewinnen diese Regressionsbäume ebenfalls datengetrieben durch die Anwendung eines Genetischen Algorithmus⁵, der die aktuelle Baum-Population gemäß definierter Rekombinationsregeln und Mutationswahrscheinlichkeiten verändert. Als Selektionskriterium wird eine Edit-Distanz verwendet, die den Unterschied zwischen einer Trainingsinterpretation und der durch Anwendung des Regressionsbaumes generierten Interpretation misst.

Mitunter ist man an einem Darstellungswechsel in die andere Richtung interessiert, also der Ableitung einer notationsähnlichen Repräsentation aus einem Strom von Steuersignalen. Typisches Anwendungsszenario für solch eine Konversion ist die Live-Eingabe von Tönen in ein Notensatzsystem über ein angeschlossenes MIDI-Keyboards. Aufgrund des höheren Abstraktionsniveaus der Notationsdarstellung sind hierbei eine Reihe von nicht-trivialen Problemen zu lösen, von denen im folgenden einige skizziert werden.

- *Tonalitäts-Erkennung*

Ein auch heute noch einflussreicher Ansatz zur Detektion der Tonart aus einer gegebenen Tonfolge stammt von Krumhansl und Schmuckler [Kru90]. Dort wird die Zusammenfassung der auftretenden Tonhöhen modulo Oktavlage gewichtet mit der Tonlänge in einem

⁴<http://www.renconmusic.org>

normierten Tonhöhen-Histogramm vorgeschlagen. Aus Trainings-Daten gewinnt Krumhansl zwei Profile (eins für Moll-, eins für Dur-Tonarten), aus denen sich durch einfache Transposition insgesamt 24 Histogramme für alle Dur- und Moll-Tonarten ableiten lassen. In gleicher Weise wird ein Tonhöhen-Profil des zu untersuchenden Stückes erstellt und mit den Referenz-Histogrammen korreliert. Die Tonart, deren Profil den höchsten Korrelationskoeffizienten aufweist, wird als Tonart des fraglichen Stückes angenommen.

Temperley verändert die Werte des Dur- und Moll-Profiles aufgrund musikwissenschaftlicher Überlegungen [Tem99] und entwickelt eine stochastische Reinterpretation des Modells [Tem02].

Einen alternativen Ansatz stellt Chew in [Che01] vor. Grundlage ist eine spezielle dreidimensionale Repräsentation von Tönen, in der Tonhöhen auf einer Helix-artigen Spirale, dem *Spiral Array*, angeordnet werden. Pro Windung werden 4 Tonhöhen im Quintabstand platziert. Hieraus resultiert, dass Töne auf direkt übereinander liegenden Windungen eine große Terz bilden. Chew's Modell ermöglicht somit die Positionierung verwandter Tonhöhen in geringem örtlichen Abstand.

Auf Basis dieser Einzel-Tonhöhen-Platzierungen lässt sich ein Akkord im Schwerpunkt der gewichteten Positionen seiner Töne verorten. Die gewichtete Mittelung solcher Positionen von Tonika-, Dominant- und Subdominantakkord ermöglicht ferner die Repräsentation einer Tonart durch einen einzelnen Punkt innerhalb der Spiral-Raums.

In ähnlicher Weise können auch mit der Tonlänge gewichtete Tonhöhensequenzen auf eine dreidimensionale Position reduziert werden. Die Bestimmung der Tonart erfolgt schließlich durch einen einfachen Nearest-Neighbor-Klassifikator, der diese Position mit den Tonart-Positionen im Spiral-Raum vergleicht.

- *Tempo- und Taktschlag-Detektion*

Will man eine MIDI-Datei, die lediglich aus zeitlich fixierten Steuersignalen besteht, in eine musikalische Notation überführen, so muss man von eventuellen interpretatorischen Ungenauigkeiten abstrahieren und jedem Anschlagszeitpunkt eine Position innerhalb einer ermittelten Taktstruktur zuweisen. Dies ist eine – insbesondere bei agogisch frei gestaltetem Ausgangsmaterial – nicht-triviale Aufgabe.

Dixon gliedert das Problem in zwei Teilaspekte: Die Detektion eines bestimmenden Stück-Tempos und die Ermittlung von Taktschlägen auf Grundlage des ermittelten Grund-Tempos [Dix01].

Zunächst findet eine Reduktion des MIDI-Datenstroms auf eine Sequenz von Anschlagszeitpunkten („rhythmischen Ereignissen“) statt. Aus paarweiser Betrachtung ermittelt man eine Menge von Zeit-Intervallen, die in einem Cluster-Prozess mit adaptiver Cluster-Zahl zu Gruppen ähnlicher Zeiträume zusammengefasst werden. Anschließend findet eine Bewertung dieser Tempo-Hypothesen statt: Hierzu wird neben der Anzahl von Ereignis-Paaren im eigenen Cluster auch die Größe „benachbarter“ Cluster berücksichtigt, deren Zeitintervall sich um ein ganzzahliges Vielfaches vom betrachteten Cluster unterscheidet. Die so ermittelte Rangfolge von Tempo-Hypothesen (und damit der Zeit zwischen zwei Taktschlägen) ist Grundlage für den nächsten Schritt des Algorithmus': die Detektion der *Position* von Taktschlägen.

Hierzu wird ein Multi-Agenten-Ansatz verwendet: Für jede Tempo-Hypothese und jedes rhythmische Ereignis in einem kurzen initialen Abschnitt wird ein Agent generiert, der

die Synchronität von Anschlagszeitpunkten mit berechneten Taktschlägen auf Grundlage seiner Tempo-Hypothese und der ersten Taktschlagposition bewertet. Die so konstruierte Agentenmenge wird sukzessive mit den Daten der MIDI-Datei konfrontiert. Jeder Agent hat dabei die Möglichkeit, sein Tempo aufgrund einer Abweichung des betrachteten Ereignisses vom erwarteten Taktschlag anzupassen und bei großen Unterschieden einen neuen Agenten mit einer Alternativhypothese zu generieren. Schließlich findet eine Gesamtbewertung der Agenten statt, in die neben der initialen Gewichtung der Tempo-Hypothese der Grad der Übereinstimmung der vorhergesagten mit den beobachteten rhythmischen Events eingeht.

Dixon verfeinert diesen Algorithmus in [Dix01] um eine Berücksichtigung der perzeptuellen Intensität der Anschlagszeitpunkte in den MIDI-Daten. Er leitet diese u.a. aus der Tondauer, der Tonhöhe und der Anschlagsstärke eines Tones ab.

Aktuellere Forschung konzentriert sich eher auf die Detektion eines Tempos aus Audio-Daten. Im Wesentlichen erweitern diese Verfahren den obigen Ansatz jedoch um eine Vorverarbeitungsschritt, der durch geeignete Signal-Verarbeitungs-Algorithmen das Audio-Signal auf eine Sequenz von rhythmischen Ereignissen reduziert, in der dann nach vorgestelltem Muster Tempo und Taktschläge identifiziert werden können [ADR04, Par04, HS08].

- *Unterscheidung enharmonisch äquivalenter Alterierungen*

Da MIDI von der Wiedergabe durch einen Synthesizer mit wohltemperierter Stimmung und diskreten Tonhöhen ausgeht, unterscheidet das Format nicht zwischen enharmonisch äquivalenten Noten, also z.B. nicht zwischen den Tönen F# und Ges. Die Entscheidung für die korrekte Tonalterierung hängt vom musikalischen Kontext, insbesondere dem lokalen harmonischen Gefüge ab.

Daher besteht der erste Schritt in dem Ansatz von Stoddard et al. [SRU04] aus einer Detektion der lokalen tonalen Zentren einer gegebenen MIDI-Datei gemäß [Mer03]⁵. Mit Hilfe dieser harmonischen Analyse lässt sich das Problem modulo lokaler Tonart auf fünf binäre Entscheidungsfragen reduzieren. So ist für einen Stückabschnitt in C-Dur lediglich die Darstellung der schwarzen Tasten fraglich. Aus Trainingsdaten, die MIDI-kodierte Musik ihrer MusicXML-Notation gegenüberstellen, induzieren Stoddard et al. fünf unabhängige binäre Entscheidungsbäume, die die Wahl zwischen Höher- und Tiefer-Alterierung auf der Grundlage verschiedener Merkmale des melodischen Kontextes (wie z.B. Histogrammen der vorausgehenden und folgenden Tonhöhenklassen) determinieren.

Steuersignal und Wellenform

Die Vielfalt an Methoden zur Umwandlung von MIDI-Daten in Audio-Signale und damit hörbare Klänge ist groß. Sie reicht von einfachen Formen der additiven Synthese (Komposition einzelner Sinus-Schwingungen) und subtraktiven Klangerzeugung (Filterung eines komplexen Ausgangssignals) über populäre Sample-Player (angepasste Wiedergabe digital aufgenommener Klänge) bis hin zu sog. Physical-Modeling-Ansätzen (physikalisch-mathematische Modellierung des Klang-Generierungsprozesses, insbesondere zur authentischen Simulation natürlicher Instrumente) sowie einer Fülle von Zwischenformen.

⁵Genauer gesagt reicht die Bestimmung der Vorzeichenmenge des tonalen Zentrum aus - für die Bestimmung der korrekten Tonalterierung ist eine Unterscheidung zwischen Paralleltonarten nicht erforderlich.

Waren diese Ansätze bis in die 90er Jahre fast ausschließlich in Hardware-Synthesizern umgesetzt, so zeigt sich heute ein verstärkter Trend zur umfassenden Tonstudio-Integration im Rechner, die den vollständigen Klangerzeugungsprozess im Computer einschließt. Wegbereitend für diese Technologie war die Standardisierung einer Schnittstelle zu solchen *virtuellen Instrumenten* durch Steinberg in Form des VST-Standards⁶.

Eine detaillierte Beschreibung von Klang-Syntheseverfahren würde den Rahmen dieser Arbeit sprengen. Für eine einführende Darstellung sei auf [Roa96] verwiesen.

Die Transformation eines akustischen Signals in eine symbolische Repräsentation der erklingenden Töne (wie z.B. MIDI) ist in der MIR-Literatur unter der Bezeichnung *Audio Transkription* ein bedeutendes Forschungsthema. Hierunter versteht man die Extraktion musikalischer Parameter der analysierten Audio-Datei, insbesondere die Identifikation und zeitliche Positionierung von hörbaren Tonhöhen und -längen. Die Verfügbarkeit einer verlässlichen und effizienten Transkriptionsprozedur würde die Zurückführung der Beantwortung musikwissenschaftlicher Fragestellungen für Audiodaten auf Analysen mit Hilfe symbolischer Verfahren gestatten. Ferner würde sie eine äußerst kompakte Darstellung des Mediums zum effizienten Retrieval erlauben.

Die Aufgabenstellung der Transkription weist starke Ähnlichkeiten zur Problematik der Spracherkennung auf: In beiden Fällen geht es um eine Art „Reverse Engineering“, um die Rückgewinnung eines „Quellcodes“, der sich in einem akustischen Phänomen artikuliert.

Klapuri gibt in [Kla04] einen Überblick über Aufbau und Leistungsfähigkeit aktueller Transkriptionssysteme. Zunächst lässt sich zwischen der Transkription monophonen und polyphonen Materials unterscheiden. Sog. *Pitch-Tracking*-Verfahren, die die Tonhöhe in einem monophonen Signal auf einer kontinuierlichen Frequenzskala ermitteln, erreichen inzwischen hohe Erkennungsraten und können als zuverlässig angesehen werden.

Die Separierung eines kontinuierlichen Tonhöhen Signals in Einzeltöne und deren Quantisierung zu diskreten Skalenstufen sind jedoch nicht trivial und insbesondere für Gesang bislang nicht vollständig zufriedenstellend gelöst [VKE03]. Dies ist einer der Gründe, warum sich das zunächst viel diskutierte „Query-By-Humming“-Paradigma zur Suche in Musikdatenbanken nicht durchsetzen konnte [GLCS95, Pau03]. Clarisse et al. geben einen vergleichenden Überblick über die Leistungsfähigkeit aktueller monophoner Transkriptionssysteme [CML⁺02].

Gemäß [Kla04] lässt sich Transkription in die Teilaufgaben der Segmentierung des Signals und der Erkennung der gleichzeitig erklingenden Töne in diesem Abschnitt zerlegen. Der Segmentierungsvorgang beruht im Wesentlichen auf oben skizzierten Ideen der Taktschlag-Detektion. Zur Identifikation der Grundfrequenz eines Tones lassen sich drei verschiedene Ansätze voneinander abgrenzen:

Eine erste Gruppe analysiert das Audiosignal im *Zeitbereich* und ermittelt die Grundfrequenz durch Autokorrelation.

Eine zweite Gruppe nutzt die Zusammensetzung eines natürlichen Instrumental-Tones aus verschiedenen Partialtönen aus, deren Frequenz ganzzahlige Vielfache der Grundfrequenz bilden, und detektiert die Grundfrequenz durch Autokorrelation im *Frequenzbereich*.

Als dritte Gruppe führt Klapuri die Verwendung komplexer auditorischer Modelle zur Tonhöhenerkennung an. Diese basieren auf hörphysiologischen Forschungsergebnissen und modellieren den menschlichen Hörprozess bis zum Hörnerv durch Filterbänke und Gleichrichter, die

⁶http://www.steinberg.net/de/company/steinberg_technology/virtual_studio_technology.html

die frequenzselektive Innervation verschiedener Sinneszellen in der Gehörschnecke nachempfinden. Ein prominenter Vertreter solch eines Pitch-Detektors ist das AMPEX-System von van Immerseel et al. [IM92].

Polyphone Transkription gilt aktuell bis auf wenige Spezialfälle, in denen die maximale Anzahl polyphoner Stimmen vorgegeben oder starke Einschränkungen an die spielenden Instrumente gestellt werden, als ungelöst. Meist wird die Problemstellung durch iterative Anwendung eines monophonen Tonhöhendetektors adressiert: Nach Erkennung einer dominanten Grundfrequenz im Signalabschnitt wird der identifizierte Ton (bestehend aus Grundfrequenz und Partialtönen) aus dem Signal entfernt und einer folgenden Grundfrequenzanalyse zugeführt. Dieses Vorgehen ist insbesondere deswegen problematisch, da in westlicher Musik eine starke Überdeckung zwischen Partialtönen eines Grundtones und anderen Grundtönen in einem harmonischen Klang existiert.

1.4 Prototyp soniXplorer

Die im Rahmen dieser Arbeit entwickelte Konzeption eines Systems zur multimodalen adaptiven Exploration von Musiksammlungen wurde in verschiedenen Prototypen implementiert und evaluiert.

Hierbei wurde ein iterativer evolutionärer Softwareprozess verfolgt:

In einer ersten Ausbaustufe erfolgte eine konzentrierte prototypische Entwicklung und Erprobung von Erweiterungen einfacher visueller Explorationsumgebungen um eine Auralisationskomponente [Lü05]. Der implementierte Prototyp realisiert eine erste multimodale Umgebung und führt das Konzept der Nutzer-gesteuerten Wahrnehmungswinkel-Beschränkung ein (vergl. Abschnitt 3.4.2).

In einer zweiten Evolutionsstufe wurden Elemente dieser Implementierung zum Aufbau einer dreidimensionalen Explorationsumgebung verwendet, die die Modell-Vektor-Struktur einer selbstorganisierenden Karte zur Topologie-erhaltenden Platzierung von Stücken im Raum und zur Abgrenzung von Clustern durch suggestive Terrain-Höhenzüge nutzt [Nei07]. Grundlage der Stück-Anordnung dieses Prototypen ist die inhaltsbasierte Distanzfunktion der Explorationsumgebung MusicMiner [MUNS05, MUTL06].

Die systematische Evaluation dieser Java-basierten Implementierung offenbarte jedoch u.a. die folgenden Defizite:

- *Rudimentäre visuelle Darstellung*

Die Präsentation von Stücken als einfache Kugeln in einer dreidimensionalen Landschaft, die durch ein unifarbenes Terrain gekennzeichnet ist, wirkt auf den Benutzer (insbesondere bei einer Vertrautheit mit aktuellen Computerspielen) zu artifiziell und erschwert eine Immersion mit der virtuellen Welt, die eine intuitivere Navigation in der Kollektion ermöglichen würde.

- *Performanz der verwendeten Game Engine*

Um aufwändige Eigenentwicklungen zum Rendering der dreidimensionalen Szene zu vermeiden, greift die Implementation auf die JavaMonkey Engine zurück. Wie sich im Verlauf der Evaluation zeigte, ist diese jedoch für eine reaktive Explorationsumgebung

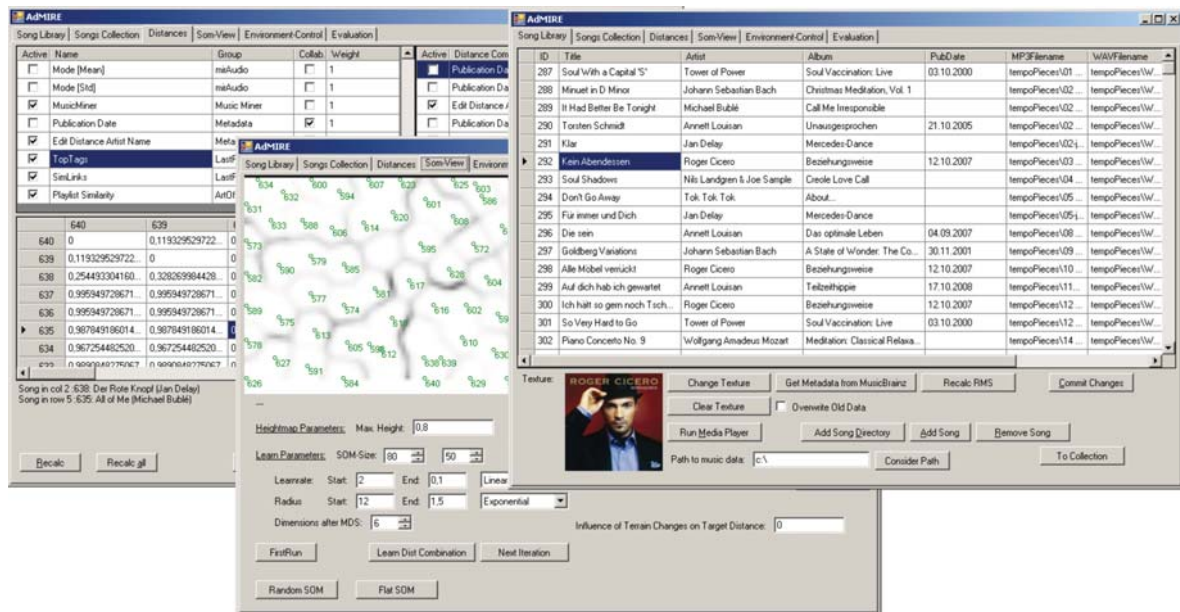


Abbildung 1.9: Screenshots des soniXplorer-Prototypen: Distanzen-Übersicht, 2D-SOM-Generierung und -Betrachtung, Song-Bibliotheksbearbeitung

nicht performant genug und erlaubt ferner nicht die Realisation gewünschter Visualisierungen (wie z.B. eine optische Korrespondenz zur Einschränkung des akustischen Wahrnehmungswinkels).

- *Qualität der Kollektions-Strukturierung*

Die integrierten inhaltsbasierten Merkmale des MusicMiners führen mitunter zu nicht nachvollziehbaren Ähnlichkeitsgruppierungen betrachteter Musikkollektionen. Die Beschränkung auf inhaltsbasierte Distanzmaße verhindert die Nutzung qualitativ hochwertiger kollaborativer Ähnlichkeitsmaße.

- *Unzulängliche Interaktionsmöglichkeiten*

Zum Aufbau einer Explorationsumgebung, in der sich Benutzer intuitiv bewegen und orientieren können, müssen dem Anwender Möglichkeiten zur Anpassung der Landschaft gegeben werden.

Diese Erfahrungen führten letztlich zur Implementierung eines dritten Prototypen, der die Bezeichnung soniXplorer trägt [LJ09]. Abbildungen 1.9 und 1.10 geben in Form von Screenshots einen kleinen Eindruck von diesem System, dessen Konzeption im Verlauf dieser Arbeit detailliert beschrieben wird.

Auf die oben genannten Probleme des vorangehenden Prototypen antwortet soniXplorer durch eine Neuentwicklung, die auf der Basis von Microsofts XNA-Gaming-Framework eine deutlich performantere und visuell ansprechendere Darstellung der Musikkollektion in einer virtuellen Landschaft ermöglicht. Ferner erlaubt die offenere Architektur des Rahmenwerks die Integration von benutzerdefinierten Vertex- und Pixel-Shadern zur Realisation spezieller grafischer Effekte (z.B. der angesprochenen Abdunklung von Szenen-Ausschnitten, die außerhalb des akustischen Wahrnehmungswinkels liegen) sowie die Ersetzung der einfachen stereophonen Standard-Audio-Wiedergabe durch eine aufwändige HRTF-basierte 3D-Spatialisierung.

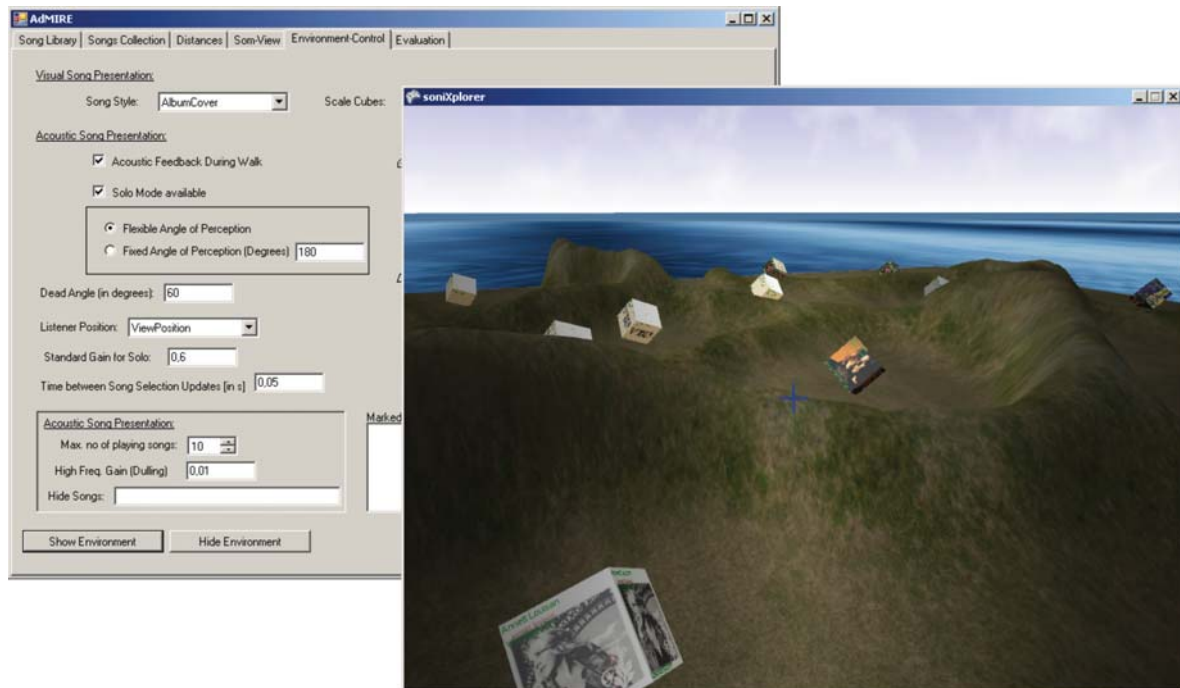


Abbildung 1.10: Screenshots des soniXplorer-Prototypen: Immersive 3D-Explorations-Umgebung und deren Parametrisierung

Darüber hinaus wurden weitere Ähnlichkeitsmaße für Musikstücke aufgenommen: Neben der Integration einer Fülle zusätzlicher inhaltsbasierter Merkmale, die unterschiedliche musikalische Aspekte (z.B. Klang oder Rhythmus) modellieren, wurden kollaborative Distanzbegriffe auf der Basis von Online-Portalen wie Last.fm oder ArtOfTheMix entwickelt. Zur Zusammenführung dieser Einzelmaße wird ein einfaches lineares Modell genutzt.

Erweiterte Interaktionsmöglichkeiten erlauben nicht nur die individuelle Gestaltung der Landschaft (z.B. durch Platzierung von Hinweisschildern oder die Ersetzung der Standard-Darstellung eines Stückes als texturierten Würfel durch suggestivere dreidimensionale Symbole), sondern ferner die automatische Adaption der Gewichte des linearen Modells, so dass das gelernte Ähnlichkeitsmaß die durch den Benutzer geänderte Strukturierung der Musikkollektion bestmöglich erklärt. Dieser Anpassungsschritt ermöglicht somit die Induktion eines personalisierten Distanzbegriffes, der eine für den Anwender nachvollziehbarere Positionierung von Titeln, die der Kollektion später hinzugefügt werden, erlaubt.

Abbildung 1.11 fasst diese Funktionalitäten in einem Prozessmodell graphisch zusammen, das den Datenfluss im soniXplorer visualisiert. Wie sich hieraus erkennen lässt, werden nach einer Kombination verschiedener Einzelmaße zur musikalischen Ähnlichkeitsbestimmung die Dimensionsreduktionsverfahren Multidimensionale Skalierung sowie selbstorganisierende Merkmalskarten angewandt, um letztlich eine multimodale virtuelle Umgebung zu schaffen, in der der Benutzer frei navigieren und die präsentierte Musikkollektion explorieren kann. Ferner ist erkennbar, wie aus der Beobachtung der Interaktion des Benutzers mit der Umgebung eine Neu-Gewichtung der Einzel-Ähnlichkeitsmaße durchgeführt wird und sich das System somit an das Ähnlichkeitsverständnis des Anwenders anpasst.

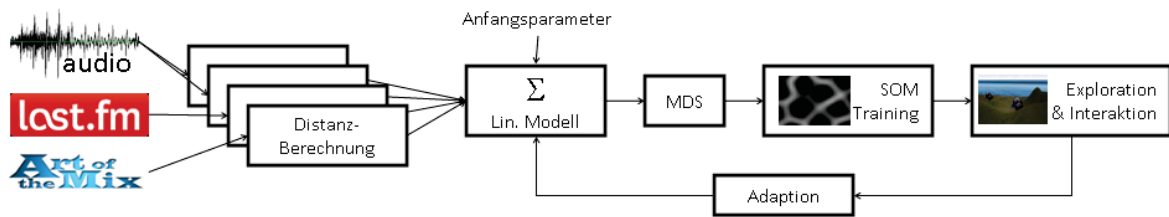


Abbildung 1.11: Kernprozesse der adaptiven multimodalen Exploration im soniXplorer

1.5 Aufbau der Arbeit

1.5.1 Struktur

Die Struktur dieser Arbeit gliedert sich in sechs Kapitel und folgt in weiten Teilen dem Prozess-Diagramm aus Abbildung 1.11:

Im ersten Kapitel wurde die Aufgabenstellung dieser Arbeit zu Grund liegende Aufgabenstellung beschrieben und motiviert. Ferner erfolgte eine kurze Einordnung in das Forschungsfeld des Music Information Retrieval sowie eine übersichtsartige Beschreibung des entwickelten Prototypen soniXplorer.

Es schließt sich in Kapitel zwei eine ausführliche Diskussion von Techniken zur Quantifizierung der Ähnlichkeit von Musik an, die für die Entwicklung einer strukturierten Explorationsumgebung zentral sind. Hierzu werden eine Fülle von aus der Literatur bekannten Ansätzen klassifiziert, einheitlich dargestellt, diskutiert und bewertet. Dieser Teil der Arbeit schließt mit einer Beschreibung der im soniXplorer letztlich eingesetzten Ähnlichkeitsmaße.

Das anschließende Kapitel stellt die multimodale Explorationsumgebung selbst in den Mittelpunkt. Nach einem ausführlichen Überblick über verwandte Arbeiten zur strukturierten Darstellung von Musikkollektionen werden die in dieser Arbeit eingesetzten Dimensionsreduktionsverfahren (Multidimensionale Skalierung und selbstorganisierende Merkmalskarten) erläutert. Daran schließt sich die Beschreibung der Visualisierung und Auralisierung der so erzeugten Kollektionsstrukturierung an.

Ein zentraler Bestandteil dieser Arbeit ist die Untersuchung der Frage, wie sich aus Veränderungen, die der Benutzer in und an der initialen Explorationslandschaft vornimmt, auf eine Adaption des Ähnlichkeitsbegriffes schließen lässt, der letztlich zur intuitiv verständlicheren personalisierten Platzierung neuer Stücke in der Umgebung verwendet werden kann. Hierzu wird in Kapitel 4 zunächst auf die implementierten Interaktionsmöglichkeiten eingegangen und anschließend erläutert, wie hieraus eine Neugewichtung der Einzel-Ähnlichkeitsmaße gelernt werden kann.

Das folgende Kapitel beschäftigt sich mit der Evaluation des entwickelten Prototypen soniXplorer. Insbesondere wird untersucht, inwiefern die Ähnlichkeits-getriebene Strukturierung einer Kollektion tatsächlich für die Orientierung des Benutzers hilfreich ist. Eine ähnliches Experiment weist die Relevanz der akustischen Wiedergabe während des Explorationsprozesses nach. In einem letzten Versuch wird schließlich die Adaptionfähigkeit des Systems an unterschiedliche Ähnlichkeitsbegriffe gezeigt.

Die Arbeit schließt mit einer Zusammenfassung und einem Ausblick auf sich anschließende zukünftige Fragestellungen.

1.5.2 Kern-Beiträge der Arbeit

Durch die systematische Entwicklung und Evaluierung des soniXplorer-Prototypen zur Explorationsunterstützung in Musikkollektionen leistet diese Arbeit für das Forschungsfeld des Music Information Retrieval die folgenden Kernbeiträge:

- *Multimodale Explorationsumgebung*

Der entwickelte Prototyp soniXplorer realisiert eine Ergänzung klassischer visueller Exploration um eine akustische Komponente, die einen intuitiven multimodalen Zugriff auf die betrachtete Musikkollektion erlaubt. Eine durch den Benutzer konfigurierbare Einschränkung des Wahrnehmungswinkels ermöglicht die Fokussierung des Hörers auf frontal positionierte Titel und entgegnet zusammen mit einem intelligenten Stück-Auswahlmechanismus so der häufig berichteten akustischen Reizüberflutung des Anwenders durch eine Vielzahl simultan erklingender Schallquellen.

In der Abschlussevaluation des soniXplorers zeigt sich klar, dass die Ergänzung visueller Präsentation durch akustische Stück-Wiedergabe die Orientierung des Benutzers in der Umgebung deutlich verbessert.

- *Adaptives Ähnlichkeitsmaß*

Das implementierte Lernverfahren zur Anpassung des Ähnlichkeitsmaßes, das der Explorationslandschaft zu Grunde liegt, an das vom Benutzer intendierte Strukturierungsprinzip auf Basis einer Analyse der Repositionierungen von Stücken in der Landschaft und Umformungen des Terrains realisiert einen adaptiven personalisierten Ähnlichkeitsbegriff und löst somit für den Anwender transparent das Problem heterogener Ordnungsprinzipien verschiedener Benutzer.

Die Leistungsfähigkeit des implementierten Adaptionverfahrens wird experimentell nachgewiesen.

- *Integration inhaltsbasierter und kollaborativer Ähnlichkeitsmaße*

Der soniXplorer kombiniert eine Vielzahl inhaltsbasierter und kollaborativer Ähnlichkeitsmaße in einem linearen Modell. Zur Ermittlung inhaltsbasierter Distanzen wird eine Reihe populärer Signalverarbeitungs-Algorithmen integriert; zur Berechnung kollaborativer Ähnlichkeit werden eigene Distanzmaße auf der Basis zugreifbarer Musikportale (Last.fm und ArtOfTheMix) entwickelt.

- *Simulation kollaborativer Ähnlichkeit durch Kombination inhaltsbasierter Maße*

In einer aufwändigen Evaluation wird die Gewichtung inhaltsbasierter Distanzfunktionen zur weitestmöglichen Simulation eines kollaborativen Referenzmaßes geschätzt. Dies ermöglicht die näherungsweise Vorhersage eines qualitativ hochwertigen kollaborativen Ähnlichkeitswertes auf der Grundlage rein inhaltsbasierter Merkmale, wenn noch nicht hinreichend viele Informationen über den Verwendungskontext der betrachteten Titel vorliegen, um eine kollaborative Bewertung zu berechnen. Die ermittelte Kombination ist signifikant besser als jedes isolierte inhaltsbasierte Maß. Sie bildet die Basis für die initiale Parametrisierung des linearen Modells vor einer Adaption an den Benutzer.

- *Vereinheitlichende Darstellung und Diskussion von Ähnlichkeitsmaßen*

In einem umfangreichen Literaturüberblick werden eine Vielzahl verschiedener populärer Verfahren zur Quantifizierung der Distanz zwischen zwei gegebenen Musikstücken

in einem einheitlichen Rahmenwerk dargestellt und vergleichend gegenübergestellt. Die untersuchten Algorithmen werden ferner hinsichtlich verschiedener Kriterien diskutiert und bewertet.

Kapitel 2

Ähnlichkeitsmaße

Eine Quantifizierung musikalischer Ähnlichkeit ist essenzielles Herzstück vieler Aufgaben des Music Information Retrieval.

Hierzu gehören nicht nur Systeme zur gezielten Suche, die zu einer gegebenen Anfrage eine nach Ähnlichkeit sortierte Liste von Stücken aus einer indizierten Musikkollektion zurückgeben (z.B. MuscleFish [WBKW96]), sondern auch Empfehlungssysteme und personalisierte Radio-Stationen, die auf der Grundlage eines erlernten Benutzerprofils individuell zugeschnittene Vorschläge z.B. in Form von Playlists generieren, die zu vom Anwender favorisierter Musik ähnlich und von ihm abgelehnter Musik unähnlich sind.

Auch für eine Explorationsumgebung ist ein für den Benutzer nachvollziehbares Ähnlichkeitsmaß unverzichtbar, auf deren Basis Musik strukturiert dargestellt und navigierbar angeordnet werden kann.

Musikalische Ähnlichkeit kann auf verschiedenen Niveaus definiert werden: Online-Portale wie Amazon konzentrieren sich meist auf die Verknüpfung von Verkaufseinheiten, was in der Regel auf einen Ähnlichkeitsbegriff zwischen *Alben* hinausläuft. Systeme, die textuelle Kontextinformationen zur Musik wie Alben-Kritiken oder Künstler-Biographien nutzen, induzieren hieraus meist Distanzen auf dem höheren Aggregationsniveau des *Künstlers*. Inhaltsbasierte Ähnlichkeitsmaße, deren Grundlage das Audiosignal selbst ist, können am natürlichsten auf *Stück*-Ebene definiert werden.

Diese Arbeit konzentriert sich primär auf letztere Aggregationsstufe. Zum einen erlauben aktuelle Download-Portale meist den Erwerb einzelner Songs, zum anderen weisen Titel eines Albums und erst recht Alben eines Künstlers eine mitunter große Heterogenität auf, die eine zusammenfassende Ähnlichkeitsberechnung nicht sinnvoll erscheinen lässt. Die Fokussierung auf ein Stück-basiertes Ähnlichkeitsmaß schließt die Integration von Zusammenhängen auf größerer Aggregationsstufe jedoch nicht aus.

Musikalische Ähnlichkeit ist ein hoch facettenreicher Begriff, dessen konkretes Verständnis nicht nur von Benutzer zu Benutzer variieren, sondern auch von der betrachteten Kollektion abhängen kann.

Die Liste möglicher ähnlichkeitsbestimmender Einflussfaktoren ist nahezu beliebig lang: Zu den dominanten Kriterien gehören sowohl klangliche Merkmale wie Instrumentierung oder die Unterscheidung zwischen männlichem und weiblichem Lead-Gesang als auch rhythmische Merkmale wie Tempo oder die Stärke eines dominanten Beats. Hinzu kommen Kontextinformationen wie das Jahr der Veröffentlichung, Herkunft der Band-Mitglieder oder ganz

individuelle Erfahrungen des Benutzers, die sich nicht aus dem Audiosignal selbst ableiten lassen.

In der Literatur wird die Definition eines Ähnlichkeitsmaßes zwischen zwei Stücken häufig auf den Vergleich ihrer *Genres* reduziert, für deren Bestimmung ein Klassifikator erlernt wird [TC02, XMS⁺03, S XK04, AP03, WC04, LOL03]. So, wie die Festlegung eines objektiven Ähnlichkeitsmaßes problematisch ist, so ist jedoch auch die Definition einer allgemein akzeptierten Genre-Hierarchie bislang nicht gelungen [SZM06, SCBG08].

Aufgrund der zentralen Bedeutung einer automatischen Distanzberechnung zwischen Stücken sind verschiedene Ansätze zur Ähnlichkeitsbestimmung mehrfach gegenübergestellt und evaluiert worden [BLEW04, MS08a, LEB03, PDW03b, SK05]. In einer der umfangreichsten Studien, die sich allerdings auf Ähnlichkeit auf Künstler-Ebene beschränkt, zeigen Berenzweig et al. die Überlegenheit von kollaborativen und Experten-basierten Ähnlichkeitsmaßen, die über eine reine Signalverarbeitung des Musikstückes hinausgehen und kulturelle Zusammenhänge abbilden können [BLEW04]. Im Vergleich so berechneter Stück-Rankings mit Nutzerfragebögen zeigt sich jedoch auch hier die Subjektivität musikalischer Ähnlichkeit, die nach einem benutzeradaptiven Ähnlichkeitsbegriff verlangt. Ferner weisen kollaborative Verfahren die als *new item problem* oder *cold start problem* bekannte Schwäche auf, bislang unbekannte Objekte, die noch von keinem Benutzer betrachtet wurden, bei der Generierung von Empfehlungen außer Betracht zu lassen. Anstatt dessen verstärkt die Empfehlung bereits populärer Objekte deren Popularität weiter (*popularity bias*). Dies macht die Entwicklung von hybriden Systemen attraktiv, die kollaborative Ähnlichkeiten mit inhaltsbasierten Informationen kombinieren.

Die folgenden Unterkapitel diskutieren verschiedene Ansätze der Ähnlichkeitsberechnung:

Abschnitt 2.1 stellt Signalverarbeitungsalgorithmen zur Extraktion höherwertiger perzeptuell relevanter Merkmale vor, die zur Modellierung eines Stückes und letztlich zur Berechnung *inhaltsbasierter* Ähnlichkeit verwendet werden.

Alternativ lässt sich die im Web veröffentlichte Meinung von *Musik-Experten* (z.B. Musik-Redakteuren) zur Ähnlichkeitsbestimmung heranziehen (Abschnitt 2.2).

Abschnitt 2.3 beschreibt die Anwendung von Verfahren der kollaborativen Filterung zur Herleitung eines Ähnlichkeitsbegriffes, der aus der Beobachtung von Kaufhistorien oder dem gemeinsamen Auftreten von Stücken auf Wiedergabelisten abgeleitet wird.

Im daran anschließenden Unterkapitel werden Algorithmen des klassischen Information Retrieval vorgestellt, die durch eine textuelle Analyse von beschreibenden Kontext-Informationen (wie etwa Rezensionen oder Künstler-Biographien) Ähnlichkeiten zwischen Musikstücken bestimmen. Dies umfasst die Beschreibung einer umfangreichen Evaluation verschiedener Retrieval-Verfahren und Datenquellen, die als Teil dieser Arbeit durchgeführt wurde.

Das folgende Unterkapitel 2.6 beschreibt Verfahren zur Kombination von unterschiedlichen Ähnlichkeitsaspekten zu einem Benutzer-spezifischen Ähnlichkeitsbegriff.

In Abschnitt 2.7 wird die Beschreibung eines im Rahmen dieser Arbeit entwickelten Verfahrens zur Simulation kollaborativer Ähnlichkeit durch rein inhaltsbasierte Merkmale vorgestellt und diskutiert, das die Qualität kollaborativer Maße mit der uneingeschränkten Anwendbarkeit inhaltsbasierter Ähnlichkeiten kombiniert.

Das Kapitel schließt mit einer Diskussion der im *soniXplorer* implementierten Einzel-Distanzmaße. Dies umfasst die detaillierte Beschreibung eigens entwickelter kollaborativer und Metadaten-basierter Ähnlichkeits-Quantifizierungen sowie die Zusammenstellung zuvor erläuteter inhaltsbasierter Distanzmaße, die im *soniXplorer* genutzt werden.

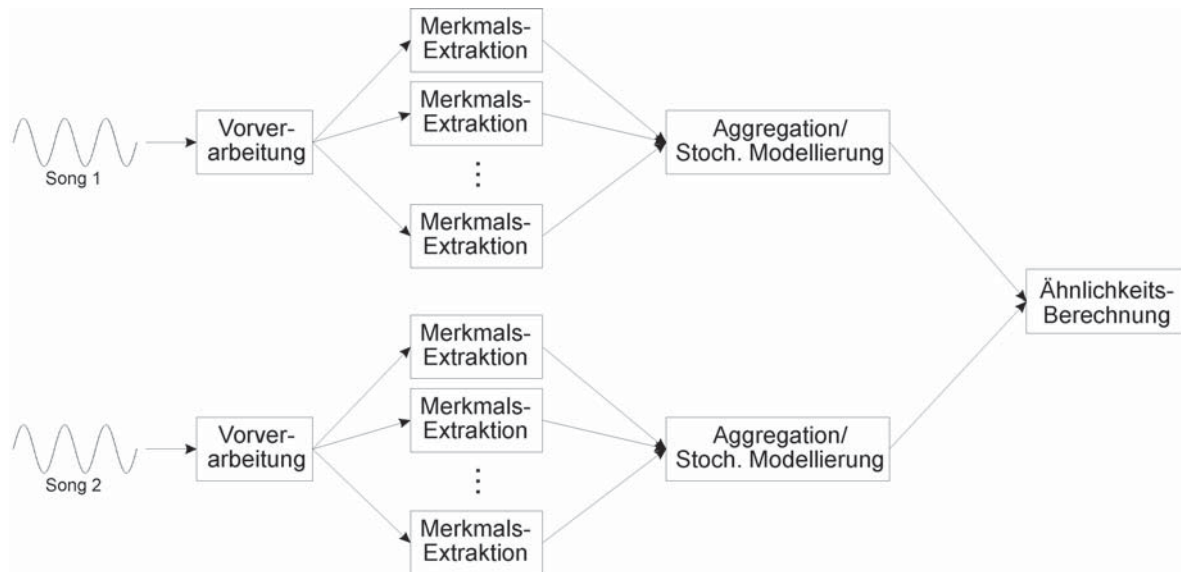


Abbildung 2.1: Prozessmodell zur Berechnung Klang-basierter Ähnlichkeit

2.1 Inhaltsbasierte Ähnlichkeit

Inhaltsbasierte Maße quantifizieren die Ähnlichkeit zwischen zwei gegebenen Stücke allein auf der Grundlage ihrer Wellenform (ihres *Inhaltes*). Damit sind sie stets anwendbar, können jedoch prinzipiell keine möglicherweise ähnlichkeitsbestimmenden Kontext-Informationen (wie z.B. Überschneidungen in der personellen Zusammensetzung der Künstlergruppen) berücksichtigen, die sich nicht aus dem Signal ableiten lassen. Ferner ist die Modellierung höherwertiger und perzeptuell wichtiger Ähnlichkeitsfaktoren auf der Basis einer Wellenform ein schwieriges informatisches Problem, das nur in Ansätzen gelöst ist [Bau05a].

Inhaltsbasierte Ähnlichkeitsmaße lassen sich grob in zwei Gruppen klassifizieren: Verfahren, die den Klang (das *Timbre*) von Musikstücken in den Vordergrund stellen und Algorithmen, die die rhythmische Struktur von Songs erfassen. Die folgenden Abschnitte beschreiben diese beiden Ansätze detaillierter.

2.1.1 Klang-basierte Merkmale

Klangliche Ähnlichkeit wird in der Literatur häufig als dominantes Kriterium zur Gruppierung von Stücken (insbesondere in Genre-Klassen) angeführt [MS08a, AF04]. Trotz der Fülle von Ansätzen zur Modellierung und zum Vergleich des Timbres zweier Stücke [Bau03, BLEW04, Foo97, BP03, HAE03, KS03, LS01, PDW03a, MUTL06] lässt sich ein recht einheitliches Prozessmodell der Verfahren identifizieren, das in Abbildung 2.1 dargestellt ist.

Nach einer Vorverarbeitung wird eine Zeitreihe Klang-relevanter Merkmale extrahiert, die im Anschluss aggregiert oder zu einem stochastischen Modell zusammengefasst wird. Schließlich quantifiziert ein Distanzmaß die Ähnlichkeit dieser Stück-Parametrisierung.

Die folgenden Unterabschnitte beschreiben typische Vorgehensweisen publizierter Verfahren in diesen Einzelphasen.

Vorverarbeitung

Zur Beschreibung der klanglichen Eigenschaften eines Stückes wird seine Wellenform durch eine Fourier-Transformation (vergl. Abschnitt 1.3.3) in eine spektrale Darstellung transformiert. Hierzu wird das Sampleraten-reduzierte Mono-Signal in kurze überlappende Fenster zerschnitten, innerhalb derer man den Klang als hinreichend stationär annimmt. Zur Vermeidung von Randeffekten multipliziert man anschließend mit einer geeigneten Fenster-Funktion. Dies lässt sich wie folgt formalisieren:

Gegeben seien ein diskretes Signal $\vec{x} = (x[0], \dots, x[N-1])^T$ mit einer Länge von N Samples, eine Fenstergröße L , eine Schrittweite H (typischerweise gilt $H = \frac{L}{2}$) sowie eine diskrete Fensterfunktion $\vec{w} = (w[0], \dots, w[L-1])$.

Dann ist der t .te gefensterte Signalausschnitt aus $\vec{x}$:

$$\vec{x}_t = (x[t \cdot H] \cdot w[0], \dots, x[t \cdot H + L - 1] \cdot w[L - 1])^T$$

Gemäß Abschnitt 1.3.3 bezeichne $F_t[k]$ den k .ten von $K = \frac{L}{2}$ unabhängigen komplexen Fourier-Koeffizienten von $\vec{x}_t$, $A_t[k] = |F_t[k]|$ den Koeffizienten des Amplituden-Spektrums und $N_t[k] = \frac{A_t[k]}{\sum_{k'=0}^{K-1} A_t[k']}$ dessen normierte Darstellung.

Fenster-basierte Merkmalsextraktion

Für jedes dieser Fenster werden im nächsten Schritt gemäß Abbildung 2.1 Merkmale zur Klangmodellierung extrahiert.

Eine ausführliche Beschreibung der in der Literatur bekannten und verwendeten Audio-Merkmale geht über den Fokus dieser Arbeit hinaus. Einen recht umfassenden Überblick gibt z.B. [MUTL06]. Die folgende Darstellung beschränkt sich daher auf die Vorstellung der populärsten Merkmale.

- *Spectral Centroid* [TC02]

Dieses Maß erfasst den *Schwerpunkt des Betragsspektrums*. Hell klingende Signale besitzen aufgrund ihres höheren hochfrequenten Signalanteils einen höheren Spectral Centroid als dumpfe, niederfrequente Signale.

$$C_t = \frac{\sum_{k=0}^{K-1} A_t[k] \cdot k}{\sum_{k=0}^{K-1} A_t[k]}$$

- *Zero Crossing Rate* [TC02]

Ähnlich zum Spectral Centroid quantifiziert die Rate der Nulldurchgänge im Zeitbereich den Anteil hoher Frequenzen.

$$ZCR_t = \sum_{i=1}^{L-1} |\operatorname{sgn}(x_t[i]) - \operatorname{sgn}(x_t[i-1])|$$

- *Spectral Rolloff* [TC02, PPW05]

Dieses Maß verwendet eine alternative Methode zur Quantifizierung des Anteils hoher Frequenzen im Signal: Beginnend mit der tiefsten Frequenz werden so lange alle Fourierkoeffizienten aufsummiert, bis ein festgelegter Anteil der Gesamt-Fourierkoeffizienten-Summe überschritten ist. Die Frequenz an diesem Punkt wird als *spectral rolloff*-Frequenz bezeichnet.

$$SR_{t,f} = \underset{\tilde{k}}{\operatorname{argmax}} \sum_{k=0}^{\tilde{k}} A_t[k] \leq f \cdot \sum_{k=0}^{K-1} A_t[k]$$

Tzanetakis et al. bestimmen die Grenzfrequenz für $f = 85\%$ der Fourierkoeffizienten-Summe [TC02], Pohle et al. summieren bis zum Erreichen der $f = 95\%$ -Grenze auf [PPW05].

- *Brightness* [Jus00]

In ähnlicher Weise lässt sich durch den Anteil der Energie im Spektrum oberhalb einer gegebenen Grenzfrequenz die Helligkeit des Klanges quantifizieren.

$$BR_{t,k_c} = \frac{\sum_{k=k_c}^{K-1} A_t[k]^2}{\sum_{k=0}^{K-1} A_t[k]^2}$$

Hierbei ist k_c der Index des Fourierkoeffizienten zur gewünschten Grenzfrequenz (üblich sind 1500 Hz).

- *Bandwidth* [LSDM01]

Die spektrale Bandbreite eines Signales (manchmal auch als *Spectral Spread* bezeichnet) lässt sich als mit den Fourier-Koeffizienten gewichtete Frequenz-Standardabweichung interpretieren.

$$BW_t^2 = \frac{\sum_{k=0}^{K-1} (k - C_t)^2 \cdot A_t[k]^2}{\sum_{k=0}^{K-1} A_t[k]^2}$$

Dichte Klänge, die das gesamte Frequenzspektrum abdecken (wie in professionell produzierter Pop-Musik) weisen in der Regel eine höhere Bandbreite auf als luftigere Jazz- oder Klassik-Aufnahmen.

- *Spectral Flatness* [JN84]

Spectral Flatness misst die Gleichmäßigkeit („Flachheit“) des Signalspektrums als Verhältnis von geometrischem zu arithmetischem Mittelwert der Fourierkoeffizienten.

$$SFL_t = \frac{\sqrt[K]{\prod_{k=0}^{K-1} A_t[k]}}{\frac{\sum_{k=0}^{K-1} A_t[k]}{K}}$$

- *Irregularity* [Jen99, KMW94]

Die Irregularität eines Spektrums ist definiert als normierte Summe der Distanzen zwischen benachbarten Fourierkoeffizienten. Ähnlich zum Spectral Flatness-Maß beschreibt sie die Gleichmäßigkeit im Spektrum, konzentriert sich jedoch auf eine lokale Umgebung. Jensen [Jen99] und Krimphoff [KMW94] stellen hierzu leicht unterschiedliche Berechnungsformeln auf.

$$IRR_{Jensen,t} = \frac{\sum_{k=0}^{K-2} (A_t[k] - A_t[k+1])^2}{\sum_{k=0}^{K-1} A_t[k]}$$

$$IRR_{Krimphoff,t} = \sum_{k=1}^{K-2} \left| A_t[k] - \frac{A_t[k-1] + A_t[k] + A_t[k+1]}{3} \right|$$

- *Spectral Flux* [TC02]

Zur Beschreibung der lokalen *zeitlichen* Veränderung des Klangspektrums berechnet man die quadrierte euklidische Distanz zwischen den normierten Spektren benachbarter Signalausschnitte.

$$SF_t = \sum_{k=0}^{K-1} (N_t[k] - N_{t-1}[k])^2$$

- *Key Clarity* [Got03, Kru90]

Dieses Merkmal beschreibt die Dominanz einer erkennbaren Tonart im Signal. Daher wird es meist für deutlich größere Signalausschnitte berechnet.

Es basiert auf der Zusammenfassung der anhand der Fourier-Koeffizienten detektierten Frequenz-Stärken zu sog. *Chromagrammen* [Got03], die die Intensität der 12 Töne der westlichen Tonskala (modulo Oktavlage) erfassen. Ein solches Chromagramm wird anschließend wie in Abschnitt 1.3.4 bereits kurz beschrieben mit typischen Mustern westlicher Tonarten verglichen [Kru90]. Man ist dabei nicht primär an der Erkennung der wahrscheinlichsten Tonart interessiert, sondern an der *Klarheit* der Entscheidung.

- *MFCCs* [Log00]

Zur Modellierung des Timbres eines Fensters verwendet man des weiteren aus der Spracherkennung bekannte Merkmale [RJ93], insbesondere Mel-Frequency Cepstral Coefficients (MFCCs) und deren diskrete Ableitungen. Hierzu gruppiert man das logarithmierte Amplitudenspektrum zu Frequenzbereichen, die äquidistant über die so genannte Mel-Skala verteilt sind und damit die wahrnehmbare Frequenzauflösung der menschlichen Gehörschnecke in kritische Bänder nachempfinden. Die so gewonnenen Filter-Koeffizienten werden schließlich mit Hilfe der Diskreten Cosinus-Transformation (als Approximation der Karhunen-Loeve-Transformation) dekorreliert (vergl. [Log00] für eine Diskussion von MFCCs in MIR-Anwendungen).

- *Psychoakustisches Spektrum* [PDW03a, PRM02]

Eine ähnliche psychoakustisch motivierte Merkmalsmenge wird von Pampalk et al. verwandt: Das Amplitudenspektrum einer Fourier-Transformation wird zunächst zur Modellierung der Frequenz-selektiven Verstärkung durch Außen- und Mittelohr gemäß Terhardt [Ter79] gewichtet. Analog zur Mel-Skala verwenden Pampalk et al. die sog.

Bark-Skala zur Gruppierung in 20 kritische Bänder [ZF99]. Im Anschluss daran werden spektrale Maskierungseffekte durch Glättungsfilter gemäß Schroeder et al. [SAH79] simuliert. So ermittelte Intensitätswerte pro kritischem Band werden zur Berücksichtigung der nicht-linearen Lautstärkenwahrnehmung des Menschen schließlich gemäß Bladon et al. zu Sone-Größen umgewandelt [RB81].

- *Spectral Contrast* [JLZ⁺02]

Jiang et al. kritisieren die Gruppierung und Zusammenfassung spektraler Komponenten zu Frequenzbändern bei der Berechnung von MFCCs und psychoakustisch motivierten Klang-Merkmalen und plädieren für die Integration eines Varianz-ähnlichen Maßes, das die Mittelwertbildung in den Frequenzbändern ergänzt. In der originären Publikation definieren Jiang et. al diesen *Spectral Contrast* auf der Grundlage von Oktav-Bändern, in späteren Merkmalsextraktoren wird das gleiche Verfahren meist auch für Mel-Bänder angewendet [AF04].

L_o bezeichne die Anzahl der Fourierkoeffizienten im Oktavband o , $(i_0, \dots, i_{L_o-1})$ eine Permutation der Fourier-Indizes zu o , so dass $(A_t[i_0], \dots, A_t[i_{L_o-1}])$ eine monoton fallende Sequenz von Koeffizienten darstellt. Zur numerisch stabileren Extremwertbestimmung in dieser Folge wird der Durchschnitt der höchsten und kleinsten Werte gebildet und logarithmiert:

$$Peak_{o,t} = \log\left(\frac{1}{\alpha L_o} \sum_{k=0}^{\alpha L_o} A_t[i_k]\right)$$

$$Valley_{o,t} = \log\left(\frac{1}{\alpha L_o} \sum_{k=L_o-1-\alpha L_o}^{L_o-1} A_t[i_k]\right)$$

Der Parameter α zur Steuerung des betrachteten Bereiches um Maximum und Minimum wird von Jiang et al. auf 0,02 festgelegt.

Der Spektrale Kontrast im Oktavband o für Fenster t berechnet sich dann als Differenz der beiden Extremwerte:

$$SC_{o,t} = Peak_{o,t} - Valley_{o,t}$$

Aggregation / stochastische Modellierung

Vorverarbeitung und Merkmalsextraktion reduzieren das Ausgangs-Audiosignal auf eine Vektor-Sequenz (ein Merkmalsvektor pro Fenster). Gemäß Abbildung 2.1 wird im nächsten Schritt aus diesen lokalen Signalcharakteristika ein globales Stück-Modell generiert, das die Datenkomplexität weiter reduziert und Songs unterschiedlicher Länge vergleichbar macht.

Hierfür sind in der Literatur verschiedene (zum Teil von der verwendeten Merkmals-Menge abhängige) Aggregations-Verfahren etabliert.

- *Empirische Momente*

Die einfachste Form der Zusammenfassung einer Sequenz von Einzelvektoren ist die Berechnung der ersten statistischen Momente (Mittelwert und Varianz, selten auch noch Schiefe und Wölbung). Der große Vorteil dieser Reduktion auf wenige Kenngrößen liegt in ihrer effizienten Berechenbarkeit und der kompakten resultierenden Modellgröße. Wie in [LS01] und [AP02] beschrieben ist die Varianz extrahierter spektraler Merkmale über

den Stückverlauf jedoch groß. Daher birgt eine Zusammenfassung zu einem einzigen Mittelwert die Gefahr einer Übereinfachung und eines Verlustes relevanter Information.

- *Zusammenfassung zu Klang-Texturen*

Tzanetakis und Cook argumentieren, dass für die Etablierung eines fundierten Klangeindrucks beim Hörer ein größerer Signalausschnitt als die bislang diskutierten wenige Millisekunden umfassenden Fenster notwendig sind [TC02]. Daher aggregieren sie die ermittelten Merkmale mittels Mittelwert- und Varianz-Berechnung zu so genannten *Klang-Texturen*, die mit einer Länge von einer Sekunde ungefähr die Dauer beschreiben sollen, die ein Mensch zur Gewinnung eines Klangeindrucks benötigt.

Die Merkmale dieses Zwischenniveaus werden schließlich über das gesamte Stück gemittelt.

- *Gauß'sche Misch-Verteilung*

Sowohl Logan und Salomon als auch Aucouturier und Pachet fassen ermittelte Einzel-fenster-Merkmale eines Stückes in einem stochastischen Modell zu einer sog. *Spektralen Signatur* zusammen [LS01, AP02]. Diese Spektrale Signatur besteht aus den Parametern einer Gauß'schen Mischverteilung über dem Merkmalsraum und modelliert die Wahrscheinlichkeitsverteilung der Merkmale somit als Überlagerung mehrerer Normalverteilungen. Formal lässt sich die Dichtefunktion dieser Verteilung wie folgt formulieren:

$$p(\vec{F}_t) = \sum_{m=1}^M c_m \cdot \mathcal{N}(\vec{F}_t, \vec{\mu}_m, \Gamma_m),$$

wobei $\vec{F}_t$ den Merkmalsvektor des k .ten Fensters und c_m , $\vec{\mu}_m$ und Γ_m Gewichtung, Mittelwert und Kovarianzmatrix der m .ten Normalverteilungskomponente mit ihrer Dichtefunktion $\mathcal{N}$ bezeichnen.

Der Ansatz von Logan und Salomon unterscheidet sich von Aucouturier und Pachet im Wesentlichen durch die Anzahl der überlagerten Normalverteilungen (Pachet: $M = 16$, Aucouturier: $M = 3$) sowie in der Methode zur Ermittlung der Verteilungsparameter: Während Logan und Salomon auf ein k-Means-Clustering-Verfahren setzen, verwenden Aucouturier und Pachet den klassischen EM-Algorithmus [Mit97].

- *Histogramme*

Pampalk et al. fassen die beschriebenen psychoakustisch motivierten Lautheitswerte pro kritischem Band in Intensitäts-Histogrammen zusammen [PDW03a]. Ein solches Histogramm ist ein zweidimensionales Diagramm mit 20 Zeilen (eine Zeile pro betrachtetem kritischem Band) und 50 Spalten, die einer Diskretisierung des Lautheits-Wertebereiches entsprechen. Die Intensität einer Zelle in diesem Diagramm beschreibt, wie häufig der betreffende Lautheitswert im jeweiligen kritischen Band erreicht oder überschritten wurde.

Mit dieser als *Spectrum Histogram* bezeichneten Modellierung repräsentieren Pampalk et al. ein Stück als $20 \cdot 50 = 1000$ -dimensionalen Vektor. Wie in [PDW03a] empirisch bestätigt, sind die Dimensionen dieses Vektors jedoch hoch korreliert: Ohne signifikanten Variabilitätsverlust ließen sich die 1000 Dimensionen durch eine Hauptkomponentenanalyse auf dominante 30 Dimensionen reduzieren.

- *Temporale Varianz*

Eine zentrale Ursache für die Datenreduktion der bislang vorgestellten Modellierungs-Ansätze liegt in der Abstraktion vom zeitlichen Verlauf und Kontext der extrahierten Merkmalsvektoren.

McKinney und Breebart weisen in [MB03] jedoch die Bedeutung dieser temporalen Variation für die Genre-Diskriminierung nach und erweitern ihr Stück-Modell daher um detektierte Modulationsintensitäten in der Merkmalsvektoren-Sequenz bei verschiedenen Frequenzen.

Ähnlichkeitsberechnung

Zur Quantifizierung der Ähnlichkeit zweier Stücke müssen schließlich die ermittelten Aggregationen oder stochastischen Modelle miteinander verglichen werden. Im Falle einer Darstellung von Stücken als (aggregierte) Merkmalsvektoren werden hierzu klassische Distanzmetriken wie Euklidische Distanz [PDW03a, TC02], Cosinus-Maß [Foo97] oder Mahalanobis-Distanz [AF04] eingesetzt.

Komplexer gestaltet sich der Vergleich von zwei Spektralen Signaturen, wie sie von Logan/Salomon und Aucouturier/Pachet in Form von Gauß'schen Mischverteilungen generiert werden. Grundlage ist die Kullback-Leibler-Divergenz zwischen zwei Wahrscheinlichkeitsdichten p und q , die als informationstheoretisches Maß die durchschnittliche Anzahl zusätzlicher Bits bei der Übertragung von Ereignissen, die p folgen, mittels einer auf q basierenden Kodierung beschreibt [DHS01].

$$KL(p, q) = \int \log \frac{p(\vec{x})}{q(\vec{x})} \cdot p(\vec{x}) d\vec{x}$$

Meist symmetriert man dieses Maß zu $KL_s(p, q) = KL(p, q) + KL(q, p)$. Trotz dieser Symmetrierung ist die KL-Divergenz in der Regel keine Metrik, da sie die Dreiecksungleichung nicht erfüllt.

Eine geschlossene Form der Kullback-Leibler-Divergenz ist für Mischverteilungen jedoch nicht formulierbar. Logan/Salomon und Aucouturier/Pachet schlagen zum Vergleich Spektraler Signaturen daher zwei alternative Ansätze vor. Jensen et al. vergleichen diese beiden Verfahren und entwickeln ein eigenes Maß.

- *Monte-Carlo Sampling* [AP02]

Aucouturier und Pachet verwenden Monte-Carlo-Sampling [Mit97] zum Vergleich der Mischverteilungen zweier Stücke: Dazu generieren sie gemäß der ersten Verteilung Merkmalsvektor-Stichproben (Samples), deren Wahrscheinlichkeit sie gemäß der zweiten Verteilung ermitteln und aufsummieren.

Dieses Verfahren garantiert ebenfalls kein symmetrisches Ähnlichkeitsmaß. Daher bilden Aucouturier und Pachet analog zur Symmetrierung der KL-Divergenz den Mittelwert der wechselseitigen Ähnlichkeit. Da aufgrund des numerischen Sampling-Verfahrens ferner nicht garantiert ist, dass die Distanz zwischen einer Spektralen Signatur zu sich selbst null ist, führen Aucouturier/Pachet abschließend eine Normierung in der Ähnlichkeitsmatrix durch.

Die Qualität eines auf Monte-Carlo-Sampling basierenden Maßes hängt maßgeblich von der Anzahl der Samples ab. Aucouturier/Pachet stellen für ihr Modell der Überlagerung von 3 Normalverteilungen erst bei 1000 Stichproben eine Stabilisierung des Distanzgrößen fest. Dies macht die Ähnlichkeitsbestimmung durch Monte-Carlo-Sampling sehr zeitintensiv.

- *Earth Mover's Distance* [LS01]

Logan und Salomon verwenden zum Vergleich Spektraler Signaturen die sog. Earth Mover's Distance [RTG98]. Diese quantifiziert den Abstand zwischen zwei Gauß'schen Mischverteilungen p und q als den Aufwand zur „Überführung“ der Verteilungsmasse von p nach q , der sich aus den minimierten „Transport-Kosten“ zwischen den verschiedenen Verteilungskomponenten von p und q zusammensetzt. Als Grunddistanz zwischen diesen Verteilungskomponenten kommt die oben beschriebene Kullback-Leibler-Divergenz zum Einsatz.

Mit Hilfe Dynamischer Programmierung lässt sich die Earth Mover's Distance recht effizient berechnen. Da bereits die KL-Divergenz die Dreiecksungleichung nicht erfüllt, stellt jedoch auch die von Logan und Salomon berechnete Distanz keine Metrik dar.

- *Normalisierte Euklidische Distanz* [JECJ07]

Jensen et al. evaluieren obige Verfahren im Rahmen einer Instrumenten-Erkennungs- und Genre-Klassifikations-Aufgabe und stellen ihnen einen eigenen Ansatz gegenüber.

Dieser normalisiert zunächst die Wahrscheinlichkeitsdichte p_i einer Spektralen Signatur:

$$p'_i(\vec{x}) = \frac{p_i(\vec{x})}{\sqrt{\int p_i^2(\vec{x}) d\vec{x}}}$$

Zwischen so normalisierten Dichten lässt sich dann der Euklidische Abstand berechnen und als Distanzmaß verwenden.

$$d_{nED}(p_1, p_2) = \int (p'_1(\vec{x}) - p'_2(\vec{x}))^2 d\vec{x}$$

Hierbei lässt sich ausnutzen, dass sich d_{nED} nach einfacher Umformung und Zusammenfassung äquivalent wie folgt beschreiben lässt:

$$d_{nED}(p_1, p_2) = 2 \left(1 - \int p'_1(\vec{x}) p'_2(\vec{x}) d\vec{x} \right)$$

Dieser Ausdruck ist geschlossen lösbar, da das Produkt von zwei Normalverteilungen wieder eine Normalverteilung einfach berechenbarer Kenngrößen samt konstantem Faktor ist [Ahr05]. Ferner offenbart obige Formel, dass die Normalisierte Euklidische Distanz als Verallgemeinerung des Cosinus-Maßes aufgefasst werden kann.

Jensen et al. stellen in ihrer Evaluation keine signifikanten Unterschiede zwischen den genannten Ähnlichkeitsmaßen fest. Im Gegensatz zu den ersten beschriebenen Ansätzen ist die Normalisierte Euklidische Distanz jedoch eine Metrik und ermöglicht somit den Gebrauch im Kontext mehrdimensionaler Indizierungsstrukturen.

2.1.2 Dynamik- und Rhythmus-basierte Merkmale

Neben den oben beschriebenen klanglichen Merkmalen, die aus spektralen Eigenschaften des Audiosignals hergeleitet werden, ist auch die rhythmische Struktur eines Musikstückes relevant bei der Formulierung eines umfassenden Ähnlichkeitsbegriffes.

Im folgenden werden daher einige Ansätze zur Modellierung und zum Vergleich rhythmischer Charakteristika von Musik auf der Grundlage des Tonsignales vorgestellt. Ausgangspunkt hierzu ist meist die Ermittlung von Periodizitäten im Audiosignal, das heißt die Detektion von durchgängigen Pulsen unterschiedlicher Tempi, die zu einem metrischen Gefüge angeordnet werden.

- *Lautstärke und Dynamik*

Ein sehr einfaches Merkmal beschreibt den Lautstärke-Verlauf des Musikstückes und basiert auf der Berechnung der durchschnittlichen Energie des Audiosignals.

Hierzu wird zunächst wie bei den spektralen Eigenschaften (vergl. Abschnitt 2.1.1) eine Segmentierung der Samplesequenz $\vec{x}$ in überlappende Signalausschnitte durchgeführt, deren *RMS-Wert* anschließend wie folgt berechnet wird:

$$RMS_t = \sqrt{\frac{1}{L} \sum_{i=t \cdot H}^{t \cdot H + L - 1} x[i]^2}$$

Der RMS-Wert stellt eine einfache Approximation des Lautstärke-Eindrucks beim Menschen dar, berücksichtigt allerdings keine hörphysiologischen und hörpsychologischen Erkenntnisse, wie diese bei der Ermittlung des psychoakustischen Spektrums (vergl. Abschnitt 2.1.1) Beachtung finden.

Aus dem Verlauf der RMS-Werte lässt sich leicht die sog. *Low Energy Rate* berechnen, die als Anteil der Fenster mit unterdurchschnittlichem RMS-Wert definiert ist.

In [Tza02] wird die Nützlichkeit dieses Merkmals zur Sprach/Musik-Diskrimination beschrieben: Das Auftreten von Pausen im gesprochenen Audio-Signal führt zu einer deutlich niedrigeren Low Energy Rate im Vergleich zum Lautstärkeverlauf eines durchgehenden Musikstückes.

- *Beat Histograms* [TC02]

Tzanetakis nimmt in seine zur Genre-Unterscheidung verwendete Merkmalsmenge eine Reihe von Signal-Charakteristika zur rhythmischen Modellierung auf. Diese leiten sich aus dem so genannten *Beat Histogram* eines Stückes ab, das wie folgt berechnet wird:

Zunächst unterteilt man das Signal in 3s lange zur Hälfte überlappende Ausschnitte. Pro Segment wird für jedes Oktavband durch Gleichrichtung, Tiefpass-Filterung und Sampleraten-Reduktion eine Mittelwert-bereinigte Hüllkurve extrahiert. Mittels Autokorrelation berechnet man die Selbstähnlichkeit der Summe dieser Hüllkurven bei unterschiedlichen Zeitdifferenzen und quantifiziert so die Intensität eines durchgehenden Pulses bei verschiedenen Tempi. Schließlich werden die Amplituden der drei dominanten Spitzenwerte dieser Berechnung, die einem Tempo zwischen 60 bpm (=Beats per Minute) und 200 bpm entsprechen, über alle Signalausschnitte in einem Histogramm zusammengefasst. Abbildung 2.2 zeigt beispielhaft Beat-Histogramme für ein klassisches,

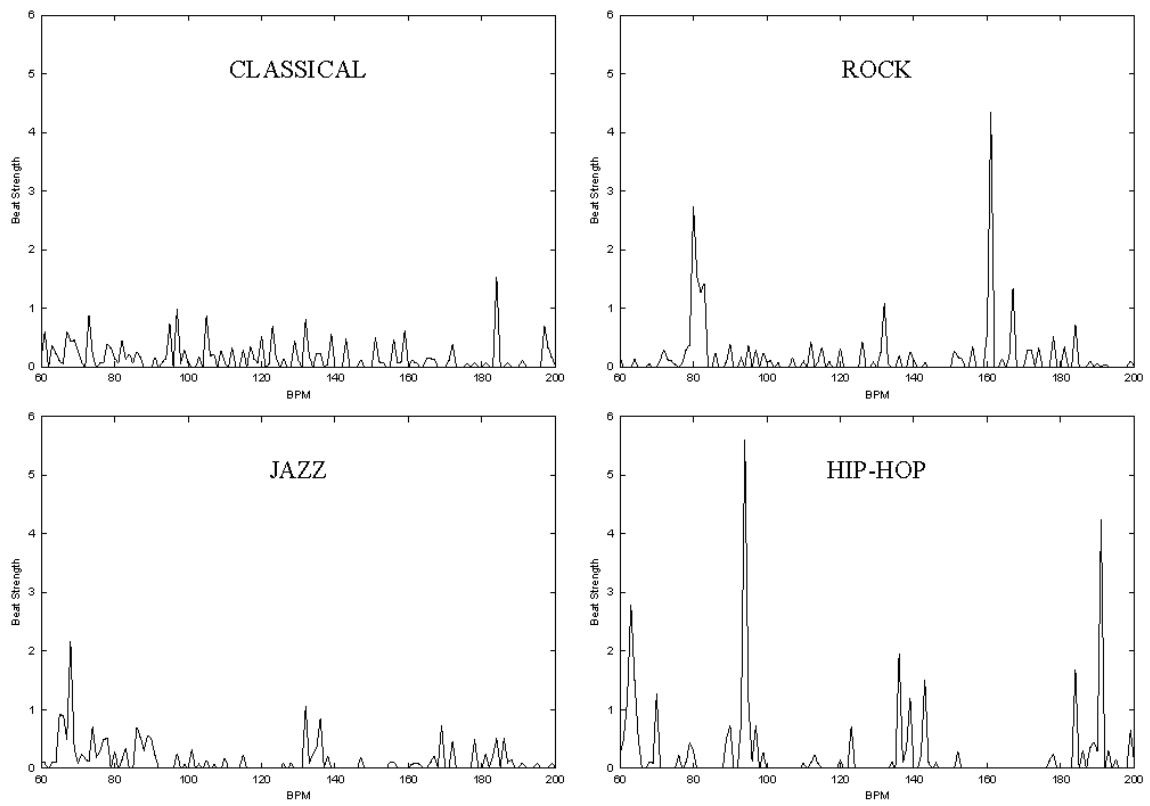


Abbildung 2.2: Beat Histograms von Stücken unterschiedlicher Genres nach [Tza02]

ein Rock-, ein Jazz- und ein Hip-Hop-Stück und illustriert deutlich das Potential dieser Repräsentation zur Genre-Diskriminierung.

Tzanetakis reduziert die Information eines Beat Histograms weiter zu sechs zusammenfassenden Merkmalen:

- Amplituden des höchsten und zweit-höchsten Spitzenwertes, normalisiert durch die Summe aller Amplituden im Histogramm
- Verhältnis der Amplitude des höchsten zum zweit-höchsten Spitzenwert
- Tempo des höchsten und zweit-höchsten Spitzenwertes in bpm
- Summe aller Histogramm-Amplituden

Neben der Extraktion des geschätzten Tempos modelliert diese Merkmalsmenge auch die Deutlichkeit eines durchgehenden Pulses. Tzanetakis kombiniert obige Merkmale in einem zusammenfassenden Vektor und schlägt den Einsatz von Standard-Metriken wie Euklidischer Distanz oder Cosinus-Maß zum Vergleich von zwei so repräsentierten Stücken vor.

- *Periodicity Histograms* [PDW03a]

Die Periodicity Histograms nach Pampalk et al. modellieren den rhythmischen Gehalt eines Stückes in ähnlicher Weise wie Tzanetakis' Beat Histograms. Ausgangspunkt ist jedoch die in Abschnitt 2.1.1 beschriebene psychoakustische Vorverarbeitung, mit der ein in 12s lange überlappende Fenster segmentiertes Signal durch eine entsprechende Filterbank in 20 Zeitreihen transformiert wird, die den Lautheitsverlauf im jeweiligen kritischen Band beschreiben. Differenzfilter und Gleichrichter erlauben die Betonung perkussiver Ereignisse im Signalverlauf. Anschließend wird mit Hilfe von Kammfiltern die Intensität verschiedener Periodizitäten (und damit Beat-Tempi) pro Band in einem Bereich zwischen 40 bpm und 240 bpm in 5 bpm-Schritten ermittelt. Pampalk et al. wenden erneut einen Gradientenfilter an und heben damit klare Spitzenwerte im Periodizitätsdiagramm hervor. Schließlich werden die Einzel-Intensitäten pro Band durch Summierung zusammengefasst. Damit lässt sich für jeden Signalausschnitt die Stärke eines wiederkehrenden Pulses bei 40 verschiedenen Tempi angeben.

Diese Fenster-genaue Information fassen Pampalk et al. wie schon bei den in Abschnitt 2.1.1 beschriebenen Spectrum Histograms zu Diagrammen zusammen, deren Zellen erfassen, wie viele Fenster für ein gegebenes Tempo ein bestimmtes Intensitätsniveau erreichen oder überschreiten. Diese *Periodicity Histograms* erlauben bei einer Diskretisierung der Periodizitäts-Intensität in 50 Werte somit die kompakte Repräsentation dominanter wiederkehrender rhythmischer Pulse in einer 40x50-Matrix.

Pampalk et al. interpretieren diese Matrix als 2000-dimensionalen Vektor. Zwei Periodicity Histograms werden mit Hilfe der Euklidischen Distanz zwischen ihren Vektordarstellungen verglichen. Wie schon bei den Spectrum Histograms berichten Pampalk et al. über die Reduzierbarkeit der Vektor-Dimensionalität auf 60 Hauptkomponenten ohne nennenswerten Variabilitätsverlust.

Wie bei Tzanetakis erlaubt die Generierung eines Periodicity Histograms prinzipiell neben der Detektion des Tempos eines Stückes auch die Beschreibung der Deutlichkeit eines durchgehenden Pulses und ermöglicht damit z.B. eine klare Abgrenzung klassischer Musikaufnahmen von HipHop-Produktionen. Ferner lässt sich aus dem Diagramm

die Unterteilung in kleinere Takteinheiten ablesen und somit z.B. eine Unterscheidung zwischen ternären und binären Rhythmen treffen.

Im Gegensatz zu Tzanetakis verzichten Pampalk et al. jedoch auf die explizite Ableitung solcher musikalisch relevanter Merkmale aus dem Periodicity Histogram. Die von ihnen durchgeführte Komposition der Diagramm-Werte in unabhängigen Dimensionen eines Vektors verwirft damit wertvolle Informationen und Zusammenhänge der Einzelkoordinaten. So ist die Distanz zwischen zwei Stücken mit den Tempi 100 bpm und 105 bpm identisch zum Abstand zwischen einem 100 bpm- und einem 200 bpm-Titel.

- *Fluctuation Patterns* [PRM02]

Ausgangspunkt der *Fluctuation Pattern*-Darstellung von Pampalk et al. ist wiederum die in Abschnitt 2.1.1 beschriebene psychoakustische Vorverarbeitung des Signals, die für jedes Fenster den Lautheitsverlauf pro kritischem Band ermittelt. Im Gegensatz zu den Kammfiltern der Periodicity Histograms wird eine einfache Fourier-Transformation zur Ermittlung der Intensität von durchlaufenden Pulsen verschiedener Tempi in einem Signalausschnitt verwendet. Ferner verzichten Pampalk et al. auf die Summation der Periodizitäts-Intensitäten über alle kritische Bänder und stellen den rhythmischen Gehalt eines Signal-Fensters schließlich in Form einer 20x60-Matrix dar, wobei eine Zeile einem kritischen Band und eine Spalte einem Tempo entspricht.

Pampalk et al. berichten über die Evaluation verschiedener komplexerer Verfahren zur Zusammenfassung dieser Fenster-basierten Merkmale zu Stück-Modellen (u.a. k-Means Clustering, Gauß'sche Mischverteilungen), entscheiden sich aber letztlich für die einfache Repräsentation eines Stückes durch den Median der Einzel-Matrizen.

Analog zu den Periodicity Histograms komponieren sie die Matrix-Werte zu einem 1200-dimensionalen Vektor, der einen Euklidischen Vergleich ermöglicht. Die Unabhängigkeit dieser Vektor-Dimensionen führt jedoch zu den bereits für die Periodicity Histograms beschriebenen Informationsverlusten.

- *Beat Spectrums* [FU01, FCN02]

Foote schlägt eine mit den Beat Histograms verwandte Repräsentation der rhythmischen Struktur in so genannten *Beat Spectrums* vor, deren Berechnung aus drei Kern-Schritten besteht:

Zunächst wird das Audiosignal in eine Sequenz von Fenster-basierten Merkmalen transformiert. Foote greift hierzu auf spektrale Kenngrößen wie Fourier-Koeffizienten oder MFCCs zurück. Anschließend wird eine quadratische symmetrische Selbstähnlichkeitsmatrix aufgebaut, deren Zelle (i, j) die Ähnlichkeit der Merkmale von Fenster i und j darstellt. Zum skalierungsinvarianten Vergleich von zwei spektralen Koeffizienten verwendet Foote das Cosinus-Maß. Aus dieser Matrix lassen sich schließlich Periodizitäten ablesen: Die Ähnlichkeit eines Signals zu einer Verschiebung von sich selbst zeigt sich in durchgängig hohen Matrixwerten auf einer Nebendiagonale, deren Abstand von der Hauptdiagonale der Verschiebung und damit dem Zeitintervall zwischen zwei Pulsen entspricht (vergl. Abbildung 2.3).

Daher fasst Foote die Summe der Nebendiagonale bei unterschiedlichen Zeitdifferenzen in einem Beat-Spektrum zusammen, das sich als Funktion der Puls-Intensität in Abhängigkeit eines Tempos interpretieren lässt. Da diese Funktion einen endlichen Träger

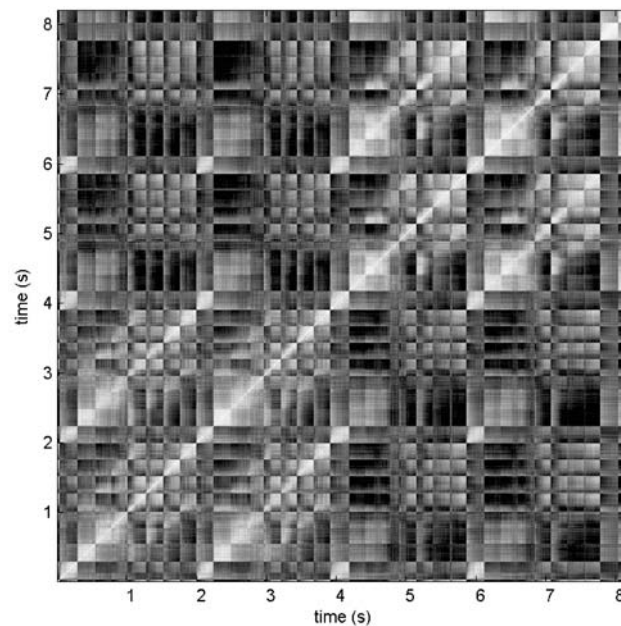


Abbildung 2.3: Selbstähnlichkeitsmatrix nach Foote [FU01]

hat, lassen sich ihre Werte als Vektorkoordinaten in einem Euklidischen Raum auffassen und entsprechend vergleichen.

- *Tempoinvarianter Rhythmusvergleich* [PK02]

Die meisten der oben beschriebenen Verfahren fassen das Tempo eines Stückes als wichtiges rhythmisches Differenzierungsmerkmal auf. Paulus und Klapuri dagegen argumentieren, dass der rhythmische Charakter („Groove“) eines Stückes in Grenzen unabhängig vom Tempo ist und schlagen daher ein Verfahren zum Vergleich von Rhythmusmustern zwischen aneinander ausgerichteten Takten vor.

Voraussetzung für diese Ausrichtung ist die Detektion von Taktgrenzen. Zur Identifikation von dominanten Periodizitäten wird ähnlich den bereits vorgestellten Verfahren eine Autokorrelation des Lautheitsverlaufs in unterschiedlichen Frequenzbänder berechnet. Ein stochastisches Modell setzt die erkannten Periodizitäten zueinander in musikalisch sinnvolle Beziehung und generiert so eine metrische Hierarchie, deren oberste Ebene dem Takt entspricht.

Dies ermöglicht eine Segmentierung des Signals in Taktabschnitte, die paarweise mit den Takten eines anderen Stückes verglichen werden können. Paulus und Klapuri erweitern die zur Takterkennung verwendete Fenster-basierte Merkmalsmenge um spektrale Kennwerte und repräsentieren den Takt eines Stückes damit als Vektor-Sequenz variabler Länge.

Mittels Dynamic Time Warping werden zwei solcher Sequenzen unter Minimierung einer Kostenfunktion aneinander ausgerichtet, die auf der Summe der Euklidischen Distanz zueinander zugeordneter Merkmalsausprägungen basiert.

Der Wert dieser Kostenfunktion ist nach Paulus und Klapuri schließlich Grundlage für

die Bemessung der rhythmische Ähnlichkeit zwischen zwei Takten. Als Ähnlichkeit zwischen zwei Stücken wird die maximale Ähnlichkeit paarweiser Taktvergleiche verwendet.

- *Notenwert-bezogene Periodizitäts-Intensität* [DPW03]

Auch Dixon et al. modellieren die rhythmische Struktur eines Stückes als Intensität von Periodizitäten in Relation zum Tempo. Anwendungskontext ihrer Merkmalsextraktion ist die Klassifikation von Audiodaten in unterschiedliche Tanzmusik-Stile. Diese Aufgabenstellung erfordert die Fokussierung auf rhythmische Charakteristika des Signals.

Zur Erkennung dominanter Periodizitäten werden zwei alternative Verfahren gegenübergestellt: Der erste Algorithmus basiert auf der Identifikation des Anfangszeitpunkts von Tönen (engl. „Onsets“). In einem iterativen Cluster-Verfahren werden die zeitlichen Abstände zwischen diesen Ereignissen (engl. „Inter-Onset-Intervals“ (IOI)) gruppiert. Das Gewicht eines so ermittelten Clusters beschreibt die Intensität einer Periodizität der Länge des Cluster-Zentrums im Signal.

Alternativ wird zur Ermittlung von Periodizitäten eine Autokorrelationsfunktion analog zum Verfahren von Paulus und Klapuri eingesetzt.

Auch Dixon et al. ordnen die detektierten Längen dominierender Pulse im Signal zu einer metrischen Struktur an, um letztlich die Länge eines Taktes zu ermitteln. Während Paulus und Klapuri sich jedoch auf den paarweisen Vergleich so identifizierter Taktsegmente beschränken, fassen Dixon et al. Informationen über den gesamten Stückverlauf in einem Merkmalsvektor zusammen. Dieser Merkmalsvektor umfasst neben dem detektierten Tempo und der erkannten Taktart die Gewichtung von Periodizitäten, ausgedrückt in Relation zum (Tempo-invarianten) *Notenwert*. So lässt sich z.B. aus der Detektion eines deutlichen punktierten Viertel-Pulses ein Samba- von einem ChaCha-Rhythmus abgrenzen.

2.2 Experten-basierte Ähnlichkeit

Voraussetzung für die Anwendbarkeit von inhaltsbasierten Ähnlichkeitsmaßen ist lediglich das Vorliegen eines repräsentativen Teils der Audio-Signale. Die hieraus resultierende Automatisierbarkeit und Skalierbarkeit kann als großer Vorteil dieser Verfahren angesehen werden.

Naturgemäß sind inhaltsbasierte Ähnlichkeitsmaße jedoch blind für Zusammenhänge, die sich nicht aus dem Klang der Musik allein ableiten lassen, für den Benutzer aber dennoch von großer Bedeutung bei der Strukturierung seiner Musikkollektion sein können. Hierzu gehören z.B. Metadaten wie das Veröffentlichungsjahr, die Herkunft einer Band oder Verbindungen der beteiligten Musiker zueinander. Dieser „kulturelle Kontext“ erfordert die Integration weiterer Informationen in einen umfassenden Ähnlichkeitsbegriff.

Ferner ist die algorithmische Detektion höherwertiger semantischer Charakteristika von Musikstücken wie z.B. die Zuordnung einer zum Song passenden Stimmung oder Emotion bislang nur unzureichend durchführbar [HDL⁺08]. Auch hier ist ein Zurückgreifen auf (ergänzende) manuell zugeordnete Metadaten sinnvoll. In einer Evaluations-Studie von Logan et al. wurde die im allgemeinen höhere Qualität Experten-basierter Ähnlichkeitsmaße im Vergleich zu inhaltsbasierten und kollaborativen Ansätzen (siehe weiter unten) nachgewiesen [LEB03].

Kategorie	Anzahl Datenbankeinträge
Alben	1.701.365
Songs	15.154.153
Künstler	1.209.741
Alben-Rezensionen	349.108
Biographien	98.445
Musik-Stile	930
Stimmungen	184
Beziehungen zwischen Künstlern oder Alben	13.794.854

Tabelle 2.1: Umfang der AllMusic-Datenbank (nach Selbstauskunft vom 10.6.2009)

Eine mögliche Quelle solcher Zusatzinformationen sind Analysen einer kontrollierten, abgegrenzten Autorengruppe von Musikredakteuren oder -wissenschaftlern (im folgenden als „Musik-Experten“ bezeichnet).

2.2.1 AMG: Annotierter Musikkatalog

So pflegt AMG einen annotierten Musikkatalog, auf den Lizenznehmer wie Microsoft, AOL, MusicMatch oder Yahoo! zur CD- und Titel-Identifikation sowie zur Anzeige von ergänzenden Informationen in ihren Medienplayern oder Online-Shops zurückgreifen. Ein Teil der zugrunde liegenden Datenbank ist über ein öffentliches Musikportal im Web¹ zugänglich.

Die dort unter dem Markennamen *AllMusic* bereitgestellten Informationen lassen sich in vier Gruppen klassifizieren:

1. *Metadaten*: Informationen über Alben, Künstler und Titel inklusive deren Genre, Veröffentlichungsdatum, Album-Cover und Platten-Label.
2. *Beschreibender Inhalt*: Stil (als Unterkategorie von Genre), Stimmung, musikalisch aktive Jahre, Geburtsdatum und -ort sowie Herkunftsland von Alben bzw. Künstlern
3. *Beziehungen*: Explizit benannte Zusammenhänge zwischen Künstlern, z.B. ähnliche Künstler oder einflussreiche und beeinflusste Künstler
4. *Redaktioneller Inhalt*: Künstler-Biographien, Alben-Rezensionen, Stil-Beschreibungen oder Bewertungen von Künstlern und Alben

Tabelle 2.1 illustriert den Umfang der gepflegten Experteninformationen nach Selbstauskunft vom 10.6.2009.

Trotz der hohen Anzahl an Künstler-Verlinkungen ist die Matrix direkter Künstler-Ähnlichkeiten dünn besetzt und nicht notwendigerweise symmetrisch. Daher fassen Ellis et al. [EW02] eine explizite Ähnlichkeitsbeziehung als ungerichtete Kante zwischen Künstler-Knoten auf und definieren die Ähnlichkeit zweier Künstler anhand der minimalen Pfad-Länge zwischen ihren Knoten. Diese sog. Erdős-Distanz ist nicht nur symmetrisch, sondern erfüllt auch die Dreiecksungleichung und ist somit eine Metrik.

¹www.allmusic.com

2.2.2 Music Genome Project (Pandora)

Ein weiteres Online-Portal, dessen Ähnlichkeitsbegriff auf manuell von Musik-Experten ermittelten Stück-Metadaten basiert, ist Pandora². Nach Eingabe eines initialen Start-Songs oder -Künstlers identifiziert Pandora ähnliche Stücke aus seinem Katalog und stellt diese dem Benutzer in Form eines personalisierten Webradio-Streams zur Verfügung. Aufgrund gescheiterter Verhandlungen mit den Rechteinhabern der Musikindustrie in Europa ist Pandora seit Mitte 2007 nur noch aus den USA zugreifbar.

Datenbasis des Ähnlichkeitsmaßes von Pandora sind Analysen des *Music Genome Projects*, das einen Merkmalsatz von ca. 400 musikalischen Eigenschaften („Genen“) zur Beschreibung eines Titels definiert. Diese Kriterien umfassen u.a. Instrumentierung, Stimmung, Struktur, Einfluss bekannter Musikstile und ggf. Charakteristika des Textes eines Stückes. Die genaue Zusammensetzung der Attributmenge ist nicht veröffentlicht. Eine umfassende nach Merkmalstyp gruppierte Darstellung findet sich jedoch in [Wik09].

Einen beispielhaften Eindruck vermittelt die folgende Auflistung der Pandora-Charakterisierung des Dolly Parton-Stückes *Stairway to heaven*:

- country influences
- bluegrass influences
- folk influences
- a subtle use of vocal harmony
- mild rhythmic syncopation
- acoustic sonority
- demanding instrumental part writing
- intricate melodic phrasing
- thru composed melodic style
- a clear focus on recording studio production
- minor key tonality
- melodic songwriting
- a heavy twang in the vocals
- acoustic rhythm guitars

Die Attributierung eines Stückes wird von geschulten Musik-Experten durchgeführt, die in einem Trainings-Programm für die Analyse gemäß den definierten Kriterien geschult wurden. Meist reduziert Pandora die Annotation eines Albums auf die Beschreibung vier charakteristischer Stücke. Zur Sicherstellung einer Experten-übergreifenden Konsistenz werden einige Titel von verschiedenen Personen analysiert.

Jedem Attribut wird eine Ausprägungsstärke auf einer zehnstufigen Skala zugeordnet. Die so ermittelten Merkmalsvektoren werden mit Hilfe einer Genre-spezifischen gewichteten Euklidischen Distanz verglichen und unter Berücksichtigung weiterer Regeln, die sich aus dem Lizenzabkommen mit der Musikindustrie ergeben, zum Aufbau einer personalisierten Playlist herangezogen [GWSK06]. Darüber hinaus hat der Nutzer die Möglichkeit, die Empfehlungen des Systems durch positives oder negatives Feedback zum gehörten Stück zu beeinflussen. Die weitestgehende Verständlichkeit des Kriterienkataloges ermöglicht die Implementation einer Erklärungsfunktion, die dem Anwender die Grundlage einer Empfehlung erläutern kann und somit zur Transparenz der Systems beiträgt. Abbildung 2.4 zeigt einen Screenshot einer solchen Empfehlung, die von einem Nutzer wie folgt kommentiert wird:

²www.pandora.com

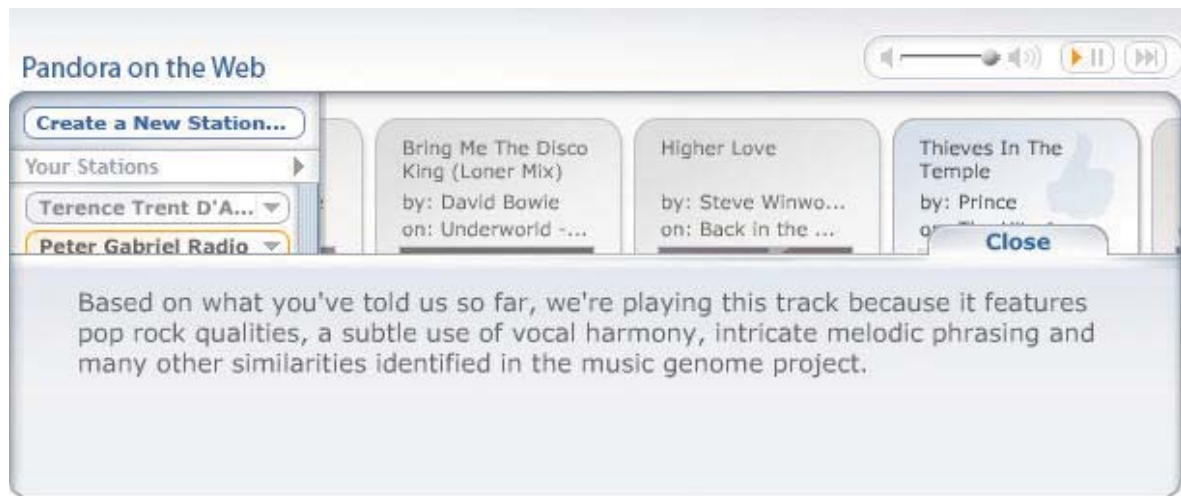


Abbildung 2.4: Screenshot der Erläuterung einer Pandora-Empfehlung [Lib07]

Pandora feels like a smart friend to me. This friend can articulate the reasons I love some of the things I love most (songs) better than I can, but only because I have told it what I like. This is one of my very favorite Prince songs and Pandora knows just why I like it so much. And I didn't know how to say it so well. Makes the technology seem very warm and reflective of my feelings and identity. It's an extension of the user, not a cold, isolating technology. I feel a part of Pandora some times. I'll bet they LOVE this. [Lib07]

Pandora listet aktuell ca. 35000 Künstler und 500000 Stücke und gibt eine aktive Nutzerzahl von 8 Millionen an. Bei einer Beschäftigung von ca. 50 Musik-Experten und einer Dauer von 20-30 Minuten pro Stück-Annotation wachse der Musik-Katalog derzeit um ca. 15000 Titel pro Monat [CL07]. Dies macht zwar eine Erfassung aktueller Main-Stream-Musik möglich, zeigt aber auch die Grenzen eines auf manueller Analyse basierenden Ansatzes auf, zumal mit steigender Mitarbeiterzahl die Gefahr inkonsistenter Beschreibungen wächst und das Vorgehen somit schlecht skaliert.

Wie bei AllMusic stehen somit einer guten Beschreibungs- und Empfehlungsqualität hohe Annotationskosten gegenüber, die eine Beschränkung auf einen (vergleichsweise kleinen) Teil des Musikmarktes erforderlich machen.

2.3 Kollaborative Ähnlichkeit

Einen in gewisser Hinsicht dazu entgegengesetzten Ansatz verfolgen Ähnlichkeitsmaße, die sich aus offenen Community-Bewertungen oder der Beobachtung des Such- und Kaufverhaltens von (potentiellen) Kunden ableiten. Hier setzt man auf eine hinreichende Fülle von Informationen, die in ihrer Qualität möglicherweise minderwertiger sind, aufgrund ihrer Quantität allerdings die Induktion dominanter Ähnlichkeitsmuster erlauben. Da solche Ähnlichkeitsmaße primär in Empfehlungssystemen Verwendung finden, die innerhalb eines gegebenen Kataloges eine Nutzer-spezifische Untermenge interessanter Objekte identifizieren, werden diese Verfahren häufig auch als *Collaborative Filtering* bezeichnet.

Grundlage eines kollaborativen Ähnlichkeitsbegriffes ist die Annahme, dass Objekte, die in gleichen Kontexten erscheinen, zueinander ähnlich sind. Im folgenden werden verschiedene solcher Kontexte betrachtet und darauf basierende Maße diskutiert.

2.3.1 Kaufhistorien und Benutzerbeobachtung

Empfehlungssysteme mit einer hinreichend großen Kundenbasis wie Amazon offerieren bei der Betrachtung einer Produktseite weitere als *verwandt* bezeichnete Angebote zum Kauf, die von anderen Benutzern des Portals zusammen mit dem aktuellen Artikel erworben wurden. Der Kontext, der eine Ähnlichkeitsbeziehung zwischen zwei Objekten bestimmt, ist hier somit die Kaufhistorie von Benutzern. Analog lässt sich die Navigationshistorie von Anwendern im Portal zur Detektion von verwandten Objekten nutzen.

Dieses Wissen kann man in einer $N \times M$ -Matrix R formalisieren, wobei N die Anzahl der Benutzer und M die Anzahl der Objekte bezeichnet. Ein von Null verschiedener Eintrag $r_{n,m}$ kennzeichnet den Erwerb des Produktes m durch Kunde n . Mitunter verwendet man anstatt einer binären Wertung den Kehrwert der Gesamt-Popularität eines Produktes, um eine Verzerrung der Empfehlung zugunsten häufig erworbener Artikel zu vermeiden [LSY03]. Ferner lässt sich eine explizite Bewertung des Artikels m durch den Benutzer in die Berechnung von $r_{n,m}$ integrieren. Ein Zeilenvektor r_n beschreibt die vom Kunden n erworbenen (und/oder bewerteten) Artikel und kann als Interessensprofil gedeutet werden.

Klassische kollaborative Empfehlungssysteme identifizieren zu einem gegebenen Benutzerprofil zunächst Kunden mit ähnlichem Kaufverhalten (*Memory-based Collaborative Filtering*, [SKKR01]). Hierzu wird z.B. die Cosinus-Distanz zwischen zwei Profil-Vektoren gebildet. Die Kaufhistorien der ähnlichsten Benutzer werden anschließend aggregiert und zu Produktempfehlungen zusammengefasst. Die Relevanz eines Produktes p für einen Benutzer u wird dabei wie folgt abgeschätzt:

$$\tilde{r}_{u,p} = \overline{R_u} + \sum_{u' \in \text{Neighbours}(u)} \text{sim}_U(u, u') (r_{u',p} - \overline{R_{u'}}),$$

wobei $\text{sim}_U(u, u')$ die oben erwähnte Ähnlichkeit zwischen Benutzer u und u' beschreibt und $\overline{R_u} = \frac{1}{M} \sum_q r_{u,q}$ die durchschnittliche Relevanz von Artikeln für Kunde u kennzeichnet.

Aufgrund des hohen Speicherbedarfs und einer Laufzeitkomplexität von $\mathcal{O}(N \cdot M)$ lassen sich solche *Nutzer-basierten Empfehlungssysteme* jedoch nicht zur direkten Generierung von Produktvorschlägen aus einem größeren Artikelstamm bei hinreichend vielen Benutzern verwenden. Linden et al. plädieren daher in ihrer Beschreibung des Empfehlungssystems von Amazon für eine *Objekt-basierten Empfehlung*, die eine Aufteilung der Berechnung in zwei Teilschritte erlaubt. Ein erster zeitintensiver Offline-Prozess, der in regelmäßigen Intervallen angestoßen wird, aktualisiert eine Produkt-Ähnlichkeitsmatrix, indem die Spaltenvektoren $r_{\cdot,p}$ und $r_{\cdot,q}$ der Artikel p und q mit Hilfe einer geeigneten Ähnlichkeitsfunktion verglichen werden.

Auf der Grundlage dieser Matrix lassen sich für ein gegebenes Benutzerprofil dann effizient die Produkte identifizieren und vorschlagen, die den bereits gekauften Artikeln des Kunden ähneln. Die Berechnung der Relevanz eines Produktes p für einen Benutzer u nach diesem Verfahren kann man demnach wie folgt formalisieren:

$$\tilde{r}_{u,p} = \frac{\sum_q \text{sim}_P(p, q) \cdot r_{u,q}}{\sum_q \text{sim}_P(p, q)}$$

Hierbei bezeichnet $\text{sim}_P(p, q)$ die zuvor berechnete Ähnlichkeit zwischen den Produkten p und q , die für diese Arbeit von primärem Interesse ist. Sie lässt sich durch Anwendung verschiedener Ähnlichkeitsmaße auf die Produktvektoren $r_{\cdot p}$ und $r_{\cdot q}$ ermitteln.

- *Cosinus-Maß*

$$\text{sim}_P(p, q) = \cos(\angle(\vec{r}_{\cdot p}, \vec{r}_{\cdot q})) = \frac{\langle \vec{r}_{\cdot p}, \vec{r}_{\cdot q} \rangle}{\|\vec{r}_{\cdot p}\|_2 \cdot \|\vec{r}_{\cdot q}\|_2}$$

Hierbei bezeichnet $\langle \vec{x}, \vec{y} \rangle$ das Skalarprodukt zwischen $\vec{x}$ und $\vec{y}$.

- *Normiertes Cosinus-Maß*

Falls die Matrix R auch explizite Kundenbewertungen von Produkten umfasst, sollte eine Verzerrung aufgrund von durchweg positivem bzw. überwiegend negativem Feedback beim Produkt-Vergleich zwischen Kundenprofilen ausgeglichen werden. Hierzu definiert man zunächst eine Menge U_{pq} , die alle Nutzer umfasst, die sowohl p als auch q gekauft bzw. bewertet haben. In der Ähnlichkeitsberechnung zwischen zwei Produkten kann man sich auf diese Nutzer beschränken. Zur Normierung der Bewertung ersetzt man den Wert $r_{u,p}$ durch seine Abweichung vom mittleren Wert des Benutzers $\bar{R}_u$:

$$\text{sim}_P(p, q) = \frac{\sum_{u \in U_{pq}} (r_{u,p} - \bar{R}_u)(r_{u,q} - \bar{R}_u)}{\sqrt{\sum_{u \in U_{pq}} (r_{u,p} - \bar{R}_u)^2} \sqrt{\sum_{u \in U_{pq}} (r_{u,q} - \bar{R}_u)^2}}$$

- *Korrelation*

Interpretiert man die Zeilen der Spaltenvektoren $r_{\cdot p}$ und $r_{\cdot q}$ zu den Artikeln p und q als Stichproben zweier Zufallsprozesse, so lässt sich die Ähnlichkeit zwischen p und q mit Hilfe des Pearson-Korrelationskoeffizienten wie folgt quantifizieren:

$$\text{sim}_P(p, q) = \frac{\sum_{u \in U_{pq}} (r_{u,p} - \bar{R}_p)(r_{u,q} - \bar{R}_q)}{\sqrt{\sum_{u \in U_{pq}} (r_{u,p} - \bar{R}_p)^2} \sqrt{\sum_{u \in U_{pq}} (r_{u,q} - \bar{R}_q)^2}},$$

wobei $\bar{R}_p$ die durchschnittliche Relevanz von p über alle Benutzer $u \in U_{pq}$ bezeichnet.

Während das Normierte Cosinus-Maß Verzerrungen durch stets positive oder negative Bewertungen unterschiedlicher Artikel eines *Benutzers* ausgleicht, gleicht das Korrelations-Maß die übermäßig positive bzw. negative Bewertungen eines *Produktes* durch unterschiedliche Kunden aus.

2.3.2 Gemeinsames Auftreten in Playlists

Viele Nutzer digitaler Musikkollektionen stellen Stücke ihrer Sammlung nach bestimmten Kriterien zu sog. Playlists zusammen. Auch ohne nähere Kenntnis der Gruppierungsprinzipien kann man die vereinfachende Annahme treffen, dass zwei Tracks, die zusammen auf einer

Playlist erscheinen, eine gewisse Ähnlichkeit aufweisen (zumindest insofern, als dass sie beide von einem Benutzer aktiv gehört wurden und somit den selben Musikgeschmack treffen).

Neben Internet-Portalen, die die explizite Veröffentlichung von Playlists (auch „Mixes“ genannt) erlauben (z.B. ArtOfTheMix³ oder Playlist.com⁴), können im gleichen Kontext gehörte Songs auch implizit durch die Beobachtung von Abspiel-Software auf dem Wiedergabegerät ermittelt werden. So bietet Last.fm für verschiedene Player Plug-Ins an, die die wiedergegebenen Stücke eines Benutzers inhaltsbasiert identifizieren („scrobblen“) und zum Aufbau eines Hörprofils an den Portal-Server übermitteln.

In der Music-Information-Retrieval-Literatur ist der Aufbau eines Playlist-basierten kollaborativen Ähnlichkeitsmaßes aufgrund der Fülle frei verfügbarer Daten recht populär und wird häufig als Grundwahrheit zur Evaluation von inhaltsbasierten Maßen verwendet (z.B. [SK05], vergl. auch Abschnitt 2.7). An die Stelle eines Benutzers und seiner Kaufhistorie im vorangehenden Abschnitt tritt nun die Playlist als zusammenfassende Klammer eines Kontextes, in dem Stücke gemeinsam auftreten. Kodiert man die Mitgliedschaft eines Songs s in einer Playlist p als binären Matrixeintrag $r_{p,s} = 1$ und setzt andernfalls $r_{p,s} = 0$, so lässt sich die Formalisierung als Ähnlichkeit zwischen Zeilenvektoren analog übertragen. Damit kann man z.B. die kollaborative Ähnlichkeit nach Stenzel et al. zwischen zwei Stücken s und t als Cosinus-Maß von Playlist-Vektoren erklären [SK05]:

$$\text{sim}_{\text{Stenzel}}(s, t) = \cos(\angle(\vec{r}_{\cdot s}, \vec{r}_{\cdot t})) = \frac{\langle \vec{r}_{\cdot s}, \vec{r}_{\cdot t} \rangle}{\|\vec{r}_{\cdot s}\|_2 \cdot \|\vec{r}_{\cdot t}\|_2}$$

Eine andere Formalisierung fasst eine Playlist p als Stückmenge auf und definiert die *Co-Occurrence* von s und t als die Anzahl der Playlists, die sowohl das Stück s als auch t enthalten:

$$\text{coc}(s, t) = |\{p | s \in p \wedge t \in p\}|$$

Analog definiert man als *Occurrence* $\text{occ}(s)$ die Anzahl der Playlists, auf denen s auftritt:

$$\text{occ}(s) = |\{p | s \in p\}|$$

Gemäß der beschriebenen Annahme ist eine hohe Co-Occurrence Indiz für eine große Ähnlichkeit. Um diese in einem Ähnlichkeitsmaß nutzen zu können, ist eine geeignete Normalisierung notwendig, die den verzerrenden Einfluss durch die unabhängige Auftretenshäufigkeit der beiden zu vergleichenden Songs korrigiert.

Keller verwendet in seinem Maß das arithmetische Mittel der Occurrences [Kel07]:

$$\text{sim}_{\text{Keller}}(s, t) = \frac{\text{coc}(s, t)}{\frac{1}{2}(\text{occ}(s) + \text{occ}(t))}$$

Um einen Wertebereich zwischen 0 und 1 sicherzustellen, wurde in dieser Arbeit im Rahmen der soniXplorer-Entwicklung eine leicht modifizierte Normierung definiert:

$$\text{sim}_{\text{Luebbbers}}(s, t) = \frac{\text{coc}(s, t)}{\min\{\text{occ}(s), \text{occ}(t)\}}$$

³www.ArtOfTheMix.org

⁴www.playlist.com

Berenzweig et al. fassen die Zugehörigkeit zu einer Playlist als Ausprägung einer Zufallsvariable auf [BLEW04]. Anwendungskontext ihrer Arbeit ist der Vergleich von Künstlern, bei geeigneten Daten ist eine Übertragung auf die Ähnlichkeitsberechnung zwischen Stücken jedoch problemlos. Ihr Ähnlichkeitsmaß basiert auf dem Vergleich der beobachteten Verbundwahrscheinlichkeit mit dem Produkt der Randverteilungen, also der erwarteten Verbundwahrscheinlichkeit unter der Annahme stochastischer Unabhängigkeit:

$$\text{sim}_{\text{Berenzweig}}(s, t) = \frac{p(s, t)}{p(s)p(t)},$$

wobei $p(s, t)$ die Wahrscheinlichkeit des gemeinsamen Auftretens von Stück s und t auf einer Playlist beschreibt und durch $\frac{\text{coc}(s, t)}{M}$ geschätzt werden kann, wenn M die Anzahl der Playlists darstellt. $p(s)$ bezeichnet die Wahrscheinlichkeit, dass eine Playlist das Stück s enthält und kann als $\frac{\text{occ}(s)}{M}$ aus den Stichproben approximiert werden. Eine informationstheoretische Rechtfertigung ihres Ansatzes ziehen Berenzweig et al. aus der Beobachtung, dass der erwartete Logarithmus von $\text{sim}_{\text{Berenzweig}}(s, t)$ der gegenseitigen Information (Mutual Information), also einer Quantifizierung des stochastischen Zusammenhanges zwischen $p(s)$ und $p(t)$, entspricht.

Berücksichtigt man, dass die Matrix R ausschließlich binäre Werte annehmen kann, so lassen sich die beschriebenen Ähnlichkeitsmaße auf eine einheitliche Notation zurückführen. Eine nähere Betrachtung offenbart, dass sie sich lediglich in der Normierung der Co-Occurrence unterscheiden.

$$\begin{aligned} \text{sim}_{\text{Stenzel}}(s, t) &= \frac{1}{\sqrt[M]{\text{occ}(s)} \sqrt[M]{\text{occ}(t)}} \cdot \text{coc}(s, t) \\ \text{sim}_{\text{Keller}}(s, t) &= \frac{2}{\text{occ}(s) + \text{occ}(t)} \cdot \text{coc}(s, t) \\ \text{sim}_{\text{Luebbers}}(s, t) &= \frac{1}{\min\{\text{occ}(s), \text{occ}(t)\}} \cdot \text{coc}(s, t) \\ \text{sim}_{\text{Berenzweig}}(s, t) &= \frac{M}{\text{occ}(s) \cdot \text{occ}(t)} \cdot \text{coc}(s, t) \end{aligned}$$

2.3.3 Tagging

Zur flexiblen Organisation von Objekten hat der Einsatz von frei zuweisbaren Schlagwörtern (engl. Tags) in der letzten Zeit an Bedeutung gewonnen. So erlaubt das Internet-Portal Flickr⁵ zur Veröffentlichung von Photos im Web die Annotation von Bildern mit beliebigen Kennzeichnungen, ein ähnliches Angebot bietet del.icio.us⁶ zur Verwaltung von Bookmarks. Im Gegensatz zur Einordnung von Objekten entlang einer festen Ontologie (wie z.B. einer vorgegebenen hierarchischen Genre-Taxonomie), vertraut man auf die Emergenz einer *Folksonomy*, die aus der wiederkehrenden Verwendung von Schlagwörtern bei Betrachtung einer hinreichend großen Nutzerbasis entsteht.

Populärstes Beispiel für ein System, das auf solch einen Bottom-Up-Ansatz zum kollaborativen Tagging von Musik setzt, ist Last.fm. Registrierte Benutzer haben die Möglichkeit, Künstler, Alben und einzelne Titel mit beliebigen Schlagwörtern zu versehen und so gemäß ihrer persönlichen Ordnungsstruktur zu organisieren. Die Aggregation der Kennzeichnungen

⁵ www.flickr.com

⁶ www.delicious.com

wobei $freq_{t,p}$ die Häufigkeit der Verschlagwortung von Stück p mit Tag t bezeichnet und dc_t die Anzahl der Stücke beschreibt, denen Tag t zugeordnet wurde. Diese Normierung reduziert das Gewicht von Tags, die vielen Stücken zugeordnet wurden und somit wenig diskriminativ sind.

Die Ähnlichkeit zweier Stücke p und q lässt sich schließlich als Cosinus zwischen den entsprechenden Spaltenvektoren der Relevanz-Matrix berechnen:

$$sim_{Tag}(p, q) = \cos(\angle(\vec{r}_p, \vec{r}_q))$$

Die Möglichkeit einer direkten Verschlagwortung ohne notwendige Einarbeitung in ein vorgegebenes Klassifikationsschema führt zu einer erhöhten Annotationsbereitschaft der Benutzer und begegnet somit erfolgreich dem Problem der Notwendigkeit einer hinreichenden Datenmenge. Der Verzicht auf jegliche Vokabular-Kontrolle birgt jedoch gleichzeitig eine Reihe von Problemen, die sich bereits am Beispiel der Tags in Abbildung 2.5 erkennen lassen. So beziehen sich die Schlagwörter „chill“, „chilled“ und „chillout“ auf ein identisches Charakterisierungskonzept, dessen Gewicht auf die verschiedenen Ausprägungen verteilt wird. Diesem Problem der Verwendung synonyme Bezeichner lässt sich nur zum Teil durch die Verwendung von Information-Retrieval- Vorverarbeitungsschritten wie einer Reduktion auf den Wortstamm ([LS07, TBL08]) begegnen. Eine weitere dazu duale Schwierigkeit stellen polysemische Begriffe dar, also Schlagwörter, deren Bedeutung (zumindest über Benutzergrenzen hinweg) mehrdeutig ist. Die Gruppe der Meinungs-Tags nach Celma und Lamere illustriert diese Problematik, ein Beispiel aus Abbildung 2.5 ist das Schlagwort „beautiful“, das in Last.fm ebenfalls zur Kennzeichnung des musikalisch eher unähnlichen Künstlers Patrick Wolf verwendet wird.

2.3.4 Diskussion

Die wachsende Popularität kollaborativer Ansätze zum Aufbau von Musik-Empfehlungssystemen und personalisierten Radiostationen ist neben einer vergleichsweise unaufwändigen technischen Realisierbarkeit auf die erreichbare hohe Qualität eines Ähnlichkeitsmaßes zwischen Stücken zurückzuführen. Diese ist u.a. durch folgende Faktoren bedingt:

- *Berücksichtigung kultureller Ähnlichkeit*

Im Gegensatz zu inhaltsbasierten Maßen, die Ähnlichkeit allein anhand des verfügbaren Tonsignales berechnen müssen, ist die Grundlage eines von Menschenhand erstellten (bzw. eines implizit aus den Kauf- und Hörgewohnheiten von Nutzern abgeleiteten) Ähnlichkeitsbegriffes vielfältiger. Diesen Vorteil teilen sich kollaborative Ähnlichkeitsmaße mit Experten-basierten Verfahren.

- *Verteilter Annotationsaufwand*

Dem Problem der schlechten Skalierbarkeit Experten-basierter Ähnlichkeitsverfahren begegnet man durch eine Verlagerung der Informationsgenerierung in die Community. Diese Arbeitsverteilung ermöglicht bei gelungener Involvierung einer hinreichenden Benutzeranzahl die Berücksichtigung einer deutlich größeren Stückmenge. Kollaborative Ähnlichkeit auf der Grundlage von Kaufhistorien oder Playlists setzt zudem keine explizite Aktivität der Anwender voraus und kann somit als konkurrenzlos kostengünstige Informationsquelle betrachtet werden.

- *Demokratischer Ähnlichkeitsbegriff*

Im Gegensatz zu Experten-basierter Ähnlichkeit, die sich auf die Meinung einer geringen Anzahl von Musikwissenden verlässt, emergiert kollaborative Ähnlichkeit aus dem Zusammenwirken und dem übereinstimmenden Urteil vieler verschiedener Benutzer. Diese Verbreiterung der Meinungsbasis führt zu einer gesteigerten Robustheit gegenüber verzerrenden Einzelwertungen.

Diesen Vorteilen kollaborativer Ähnlichkeitsmaße stehen einer Reihe konzeptueller Nachteile gegenüber:

- *Cold Start Problem*

Das gewichtigste Problem kollaborativer Ähnlichkeitmaße ist ihre Abhängigkeit von einer hinreichend großen Nutzer- bzw. Bewertungsbasis. Artikel, die neu in einen Katalog aufgenommen werden und noch von keinem Kunden gekauft oder näher betrachtet wurden, finden keinen Weg in die Menge generierter Empfehlungen („New Item Problem“).

Wie bereits erwähnt liegt in diesem Punkt die Stärke inhaltsbasierter Maße, die bei Verfügbarkeit der Klangdaten stets anwendbar sind. Zum Aufbau hybrider Empfehlungssysteme untersucht man daher, inwiefern sich mit Methoden des Maschinellen Lernens Vorhersagemodelle für kollaborative Annotationen auf der Basis inhaltsbasierter Merkmale entwickeln lassen [TBTL08, SLC07, WE04].

Aus gleicher Motivation wurde im Rahmen dieser Arbeit in aufwändigen Experimenten eine geeignete Kombination akustischer Merkmale ermittelt, die ohne den Umweg über eine generierte Schlagwortmenge einen kollaborativen Ähnlichkeitsbegriff simulieren (vergl. Abschnitt 2.7).

Die Qualität einer kollaborativen Empfehlung ist ebenso abhängig vom Wissen über den Benutzer, d.h. von der Verfügbarkeit eines hinreichend präzisen Nutzer-Profiles, das sich jedoch erst durch die Interaktion mit dem System entwickelt (z.B. durch den Kauf von Artikeln). Diese fehlende Kenntnis über die Interessen eines neuen Anwenders wird als „New User Problem“ bezeichnet. Beide Datenmangel-Probleme fasst man als sog. „Cold Start Problem“ [CL07] zusammen.

- *Early Rater Bias*

Empfehlungssysteme auf der Basis von kollaborativer Ähnlichkeit motivieren den Benutzer zur Betrachtung von Artikeln, die von „ähnlichen“ Anwendern im gleichen Kontext gekauft wurden. Sie erhöhen damit die Wahrscheinlichkeit eines Erwerbs des vorgeschlagenen Produktes, der seinerseits die Assoziation beider Artikel für kommende Empfehlungen verstärkt.

Analoges gilt für die Zusammenstellung von Playlists auf Grundlage der Playlists anderer Benutzer.

Ohne äußere Impulse ist kollaborative Ähnlichkeit in solch einem Anwendungskontext ein selbstverstärkendes Maß. Nach Celma und Lamere kommt damit den initialen Zusammenhängen, in denen ein Stück gehört oder gekauft wird, erhöhte Bedeutung zu [CL07], was die generierten Empfehlungen zu Gunsten der Meinung früher Nutzer des Artikels verzerrt („Early Rater Bias“).

- *Rauschen*

Die Annahme, dass von einem Nutzer betrachtete Artikel eines Kataloges zueinander

ähnlich sind oder dass zwei Stücke einer Playlist den gleichen Musikgeschmack treffen, sind vereinfachend und nicht notwendigerweise zutreffend. In ähnlicher Weise ist die Qualität eines kollaborativer Ähnlichkeitsbegriffs auf der Grundlage von freien Schlagwörtern durch die genannten Probleme mehrdeutiger der synonymen Begriffe gestört.

Erst durch eine hinreichend große Datenbasis heben sich wiederkehrende wertvolle Zusammenhänge von individuellem Rauschen ab und ermöglichen die Nutzung in kollaborativen Ähnlichkeitsmaßen.

- *Interessantheit / Neuigkeit*

Wie in [MS08a] herausgestellt, können kollaborative Empfehlungssysteme zwar eine hohe Präzision erreichen; wie wertvoll die generierten Vorschläge für den Benutzer sind, hängt jedoch auch von anderen Faktoren ab. Magno und Sable kritisieren die Empfehlungen von Last.fm wie folgt:

Last.fm rarely surprises you: It delivers conventional wisdom on hyperdrive, and it always seems to go for the most obvious, common-sense picks [MS08a].

Dies macht deutlich, dass zu einer Bewertung der Leistung und insbesondere des Mehrwertes eines Empfehlungssystems weitere Maße wie die *Interessantheit* oder *Neuigkeit* der erzeugten Ergebnisse („novelty“) herangezogen werden sollten (vergl. [BYRN99]).

Im Zusammenhang einer Verwendung des Ähnlichkeitsmaßes zur Strukturierung einer gegebenen Musiksammlung erscheint dieses Problem weniger relevant.

2.4 Textbasierte Ähnlichkeit

Eine weitere Quelle kontextueller Informationen, die zur Ableitung eines Ähnlichkeitsbegriffes genutzt werden kann, sind textuelle Beschreibungen der Musik z.B. in Form von Rezensionen durch Musikkritiker oder ausformulierten Nutzer-Bewertungen in Online-Portalen und MP3-Blogs [CCH06]. Bei vokaler Musik kann man des weiteren den Liedtext als Quelle textueller Ähnlichkeit nutzen [LKM04]. Grundlage der Verwendbarkeit dieser auch als *community metadata* [WL02] bezeichneten Daten ist die Annahme, dass ähnliche Beschreibungen auf eine Verwandtschaft der beschriebenen Musik hindeuten. Neben der anspruchsvollen Aufgabe einer semantischen Analyse natürlichsprachiger Texte ist man des weiteren jedoch mit dem Problem konfrontiert, dass die genannten Quellen auch eine Fülle irrelevanter Informationen beinhalten, die eine genaue Charakterisierung verrauschen [TBL08] und die Verfügbarkeit einer hinreichend großen Datenmenge erfordern.

In gewisser Hinsicht kann die Analyse textueller Beschreibungen als Vorstufe einer Tag-basierten Ähnlichkeit 2.3.3 betrachtet werden, da man das Problem gemäß dem klassischen Vektorraum-Retrieval [BYRN99] meist auf einen Vergleich zwischen Vektoren gewichteter Terme zurückführt, die aus der Beschreibung und einem betrachteten Text-Korpus ermittelt werden.

Im folgenden werden einige Ansätze zum Aufbau und zum Vergleich solcher Term-Vektoren vorgestellt.

2.4.1 Semantische Analyse von Liedtexten [LKM04]

Logan et al. untersuchen in [LKM04], inwiefern die Analyse gesungener Liedtexte zur Herleitung eines sinnvollen Ähnlichkeitsbegriffes verwendet werden kann. Um auf eine hinreichend große Datenmenge zurückgreifen zu können, konzentrieren sie sich auf Ähnlichkeit auf Künstler-Ebene und konkatenieren die Liedtexte aller Stücke eines Künstlers zu einem großen Dokument, das als textuelle Charakterisierung der Musik des Künstlers angenommen wird.

Zentraler Schritt in der Ähnlichkeitsbestimmung ist die Anwendung einer *Probabilistic Latent Semantic Analysis (PLSA)* [Hof99], die die unüberwachte Identifikation von „latenten“ Themen-Clustern innerhalb einer gegebenen Textsammlung ermöglicht. Logan et al. nutzen zum Aufbau dieser Struktur eine Zusammenstellung von über 15.000 Liedtexten.

Nach einer Stopp-Wort-Elimination erlaubt das PLSA-Modell die Abbildung eines Künstler-Dokumentes auf einen nieder-dimensionalen Vektor, dessen Komponenten die Wahrscheinlichkeit darstellen, mit der das betrachtete Dokument zum jeweiligen Themen-Cluster gehört. Logan et al. verwenden schließlich die L1-Metrik zum Vergleich von zwei so ermittelten Künstler-Modellen.

Ferner weisen sie in [LKM04] empirisch nach, dass eine Term-Reduktion auf den Wortstamm (*stemming*) sowie die Elimination von häufigen Termen, die in mehr als der Hälfte aller Themen-Modelle auftreten und somit wenig diskriminativ sind, das Ähnlichkeitsmaß (gemessen an einer aus Nutzer-Fragebögen abgeleiteten Grundwahrheit) erheblich verbessert.

Mit diesen Modifikationen ergibt sich zwar eine signifikant höhere Konsistenz mit der Grundwahrheit als bei einer zufälligen Gruppierung, man erreicht jedoch nicht das Niveau aktueller inhaltsbasierter Verfahren.

Dieses Ergebnis illustriert, dass Liedtexte nur eine bestimmte (und recht beschränkte) Dimension von Ähnlichkeit erfassen. So entspricht es wohl kaum den Erwartungen eines Benutzers, die zahlreichen Cover-Songs eines Beatles-Klassikers wie „Let it be“ in musikalisch ganz unterschiedlichen Stilen als zueinander ähnlich zu klassifizieren.

2.4.2 Textuelle Ähnlichkeit nach Whitman [WE04, WL02]

Whitman et al. stützen ihr Ähnlichkeitsmaß auf Rezensionen von Musik-Alben, die sie ebenfalls zu Dokumenten über einen Künstler zusammenfassen. Auch sie beschränken sich damit auf einen Ähnlichkeitsbegriff auf Künstler-Ebene.

Zur Datenaquise wird eine Anfrage des Musters "<Künstler>" `music review` an Google gestellt. Der Quelltext der ersten 50 zurückgegebenen Resultat-Seiten wird durch einen HTML-Parser auf den darzustellenden Text reduziert. Whitman et al. untersuchen verschiedene Alternativen der sich anschließenden Merkmalsextraktion:

- *n-Gramme*: Geordnete Wort-Sequenzen der Länge n (für $n = 1$ (Unigramme) und $n = 2$ (Bigramme))
- *Adjektive*: Ein „part-of-speech-tagger“ [Bri92] klassifiziert die Wörter eines Satzes in Wortarten. Anschließend reduziert man die Termmenge auf die identifizierten Adjektive, von denen man trotz ihrer kompakten Darstellung ein hohes Charakterisierungspotential erwartet.

- *Substantiv-Phrasen*: In ähnlicher Weise werden Substantive und sie umgebende Adverbien und Adjektive extrahiert.
- *Künstler*: Die Reduktion auf Namen anderer Künstler, die in Rezensionen eines Albums auftreten, führt letztlich auf ein Co-Occurrence-Maß (vergl. Abschnitt 2.3).

Gemäß der Evaluation von Whitman et al. [WL02] zeigen Unigramme, Bigramme sowie Substantiv-Phrasen eine im Vergleich zu den anderen Alternativen höhere Performanz.

Die so reduzierte Term-Sequenz für einen Künstler wird anschließend auf ein gewichtetes Term-Histogramm abgebildet. Hierzu verwenden Whitman et al. eine Variante des populären $tf \cdot idf$ -Schemas. Anstatt nur die Gewichte populärer Terme (d.h. von Termen mit hoher Dokumenten-Häufigkeit) zu reduzieren, plädieren sie für die ergänzende Herabsetzung des Einflusses von Termen auf die Ähnlichkeitsberechnung, die nur äußerst selten auftreten. Dies mündet in der folgenden Formalisierung der Gewichtung eines Termes t für einen Künstler a als Gauß-Funktion:

$$w_{t,a} = tf_{t,a} \cdot e^{-\frac{(dc_t - \mu)^2}{2\sigma^2}},$$

wobei $tf_{t,a}$ die Termhäufigkeit von t im Google-Resultat für Künstler a beschreibt und dc_t die Anzahl der Künstler darstellt, in deren Beschreibung t auftritt. μ und σ werden nach Whitman et al. empirisch bestimmt.

Die Modellierung der textuellen Beschreibung eines Künstlers als Vektor gewichteter Terme ermöglicht schließlich den Vergleich mit Hilfe des im Vektorraum-Retrieval üblichen Cosinus-Maßes.

In einer späteren Veröffentlichung ersetzen Whitman et al. die Suche nach beschreibenden Web-Seiten mittels einer Google-Anfrage durch den gezielten Zugriff auf die Rezensionen-Portale AMG⁷ und Pitchfork⁸ [WE04].

Der Ansatz von Baumann et al. ähnelt dem obigen Verfahren [BH03]. Irrelevante Google-Ergebnisse werden durch eine Kaskade heuristischer Filter entfernt (u.a. werden URL und Titel der zurückgegebenen Seiten untersucht).

Weitere Varianten des Verfahrens definieren eine fixe Menge von beschreibenden Termen, auf die die Titel-Modelle reduziert werden. Solche Term-Mengen entstammen z.B. der Genre-Taxonomie von AllMusic [VB04] oder einem unüberwachten Lernprozess [TTBL07, TBTL07].

2.4.3 Textuell-inhaltsbasierte Ähnlichkeit nach Knees [KPS⁺08, KPSW07a, KW07]

Das Verfahren zur Herleitung eines hybriden Ähnlichkeitsmaßes von Knees et al., das textuelle und inhaltsbasierte Ansätze kombiniert, kann als eine weitere Variante des klassischen Vektorraum-Retrievals aufgefasst werden.

Knees et al. gehen von einem abgeschlossenen Musikkatalog aus, für dessen Stücke bereits eine paarweise inhaltsbasierte Ähnlichkeit $sim_{cb}(s, s')$ errechnet wurde.

Ausgangspunkt der textuellen Analyse ist erneut die Aquis von Google-Suchergebnissen. Zum Aufbau von Titel-spezifischen Beschreibungen werden die Resultate von drei Suchanfragen kombiniert:

⁷www.allmusic.com

⁸www.pitchfork.com

- "<Künstler>" music
- "<Künstler>" "<Album>" music review
- "<Künstler>" "<Titel>" music review -lyrics

Nach Stopp-Wort-Elimination und Wortstamm-Reduktion wird die Dimension der Merkmalsraumes weiter durch einen χ^2 -Test verringert. Für jedes Stück werden hierzu die 100 gemäß dem inhaltsbasierten Maß ähnlichsten Titel a und unähnlichsten Songs b ermittelt. Anschließend wird für jeden Term seine Bedeutung für die Abgrenzung zwischen a und b quantifiziert. Hierzu verwendet man das χ^2 -Maß, das für jeden Term t die Abhängigkeit zwischen der Zugehörigkeit eines Stückes zur Menge a bzw. b und dem Vorhandensein von t in seiner Beschreibung berechnet. Die Aggregation dieser Größe über alle Titel ermöglicht schließlich eine Merkmalsreduktion auf die Terme mit der größten Beschreibungskraft.

Knees et al. verwenden das als *ltc-Variante* [SB88] bekannte Gewichtungsschema, um die Bedeutung eines Terms für ein Dokument zu quantifizieren:

$$w(t, s) = \begin{cases} (1 + \log_2 tf_{t,s}) \cdot \log_2 \frac{N}{dc_t} & : tf_{t,s} > 0 \\ 0 & : sonst \end{cases}$$

Hierbei bezeichnet $tf_{t,s}$ (wie gewohnt) die Häufigkeit des Terms t in der Beschreibung von Stück s und dc_t die Anzahl der Stücke, in deren Beschreibung t auftritt. Die Gesamtanzahl der Stücke im betrachteten Katalog ist N .

Eine weitere Anwendung erfährt das inhaltsbasierte Maß zur „Glättung“ von Termvektor-Repräsentationen zwischen inhaltsbasiert benachbarten Stücken. Hiermit entgegen Knees et al. der Gefahr einer zu geringen Datenbasis, die sich als Konsequenz aus der Konzentration auf Song-spezifische Ähnlichkeit (insbesondere bei unpopulären oder neuen Stücken) ergibt. Der Einfluss des Term-Modells n benachbarter Stücke wird durch eine Gauß-Funktion formalisiert:

$$w_{gauss_n}(t, s) = \sum_{i=0}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(i/2)^2}{2}} \cdot w(t, sim_{cb,i}(s)),$$

wobei $sim_{cb,i}(s)$ den i -ten ähnlichsten Titel zu s gemäß dem inhaltsbasierten Maß sim_{cb} beschreibt und $sim_{cb,0}(s) := s$ definiert ist.

Diese so modifizierten Term-Gewichte eines Stückes werden anschließend Cosinus-normalisiert. Dies ermöglicht einen Betrags-invarianten Vergleich zwischen zwei Song-Modellen mit Hilfe der Euklidischen Distanz.

2.4.4 Berücksichtigung von Nutzer-Feedback

Die Rückführung musikalischer Ähnlichkeit auf einen semantischen Vergleich beschreibender Texte eröffnet die Nutzung von Technologien des klassischen Information Retrieval, die eine Integration von Nutzer-Feedback in den Suchprozess und somit den Aufbau eines personalisierten Empfehlungssystems erlauben [KW07].

Unter Verwendung des Vektorraum-Modells lässt sich das Informationsbedürfnis oder das Interessensprofil eines Benutzers als Vektor von Termgewichten formalisieren. Die positive oder negative Bewertung von Suchergebnissen durch den Anwender kann dann als lineare

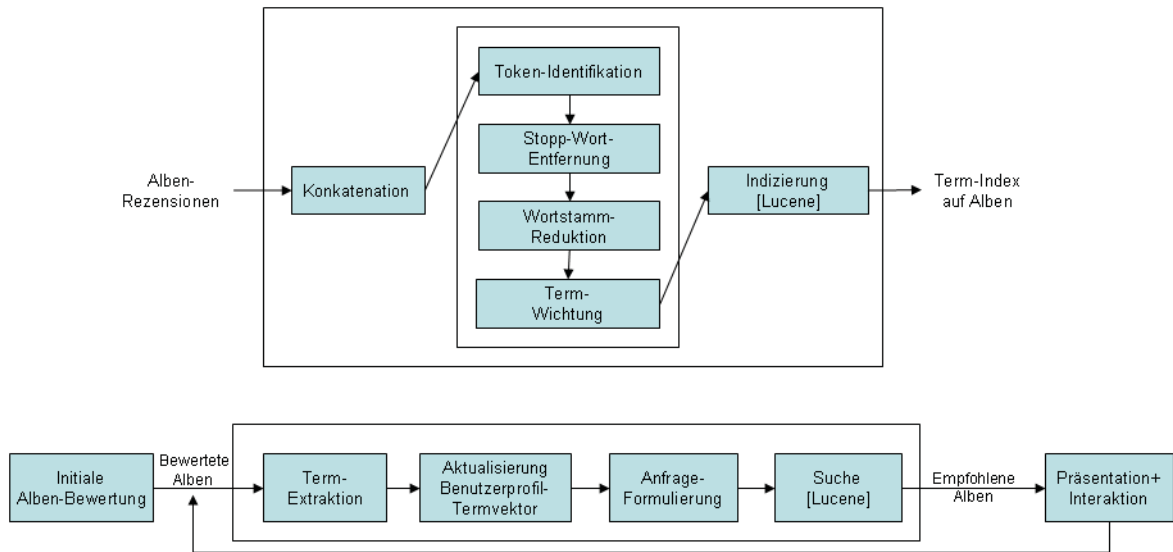


Abbildung 2.6: Information-Retrieval-Ansatz zur Rezensions-basierten Alben-Empfehlung

Verschiebung dieses Vektor entlang der Richtung der beurteilten Resultate implementiert werden:

$$\vec{q_{t+1}} = \alpha \cdot \vec{q_t} + \frac{\beta}{|D_p|} \sum_{d \in D_p} \vec{w_d} - \frac{\gamma}{|D_n|} \sum_{d \in D_n} \vec{w_d}$$

Hierbei bezeichnet $\vec{q_{t+1}}$ den aktualisierten Anfragevektor, D_p und D_n sind die Menge der positiv bzw. negativ bewerteten Suchergebnisse d mit ihren Term-Gewichts-Vektoren $\vec{w_d}$. Mit Hilfe der Parameter α, β und γ lässt sich die Stärke der Anpassung des Ausgangsvektors an das Nutzerfeedback regulieren.

Im Rahmen einer aufwändigen Evaluationsstudie, die im Kontext dieser Dissertation entstand, konnte die Überlegenheit dieses als *Rocchio Algorithmus* [Roc71] bezeichneten Ansatzes gegenüber maschinellen Lernverfahren zur Benutzerprofil-Modellierung nachgewiesen werden [Yan08].

Hierzu wurde zunächst mit Hilfe des Genre-Kataloges von Amazon.com eine heterogene Menge von 1000 Alben bestimmt, für die im Anschluss redaktionelle Produktbeschreibungen sowie Nutzerkommentare von Amazon.com, AllMusic und metacritics⁹, einem Portal, das eine Zusammenstellung der Rezensionen vielfältiger Internet-Quellen offeriert, aquiriert wurden.

Abbildung 2.6 zeigt die Einzelschritte des implementierten Prozesses zum Aufbau eines Retrieval-Indexes sowie der realisierten Feedback-Schleife: Zur Abschwächung des oben genannten *cold start*-Problems werden die Alben des Kataloges zunächst mit Hilfe eines k-Mean-Clustering-Verfahrens in disjunkte Gruppen partitioniert, deren (maximal unterschiedliche) Prototypen dem Benutzer zur initialen Bewertung vorgelegt werden. Auf der Grundlage der Rezensionen zu den bewerteten Alben wird anschließend ein gewichteter Termvektor aktualisiert, der in diesem Information-Retrieval-Ansatz die Benutzerinteressen modelliert. Dieser Anfrage-Vektor erlaubt schließlich die Verwendung der Standard-Engine *Lucene*¹⁰ zur Suche

⁹ www.metacritics.com

¹⁰ <http://lucene.apache.org>

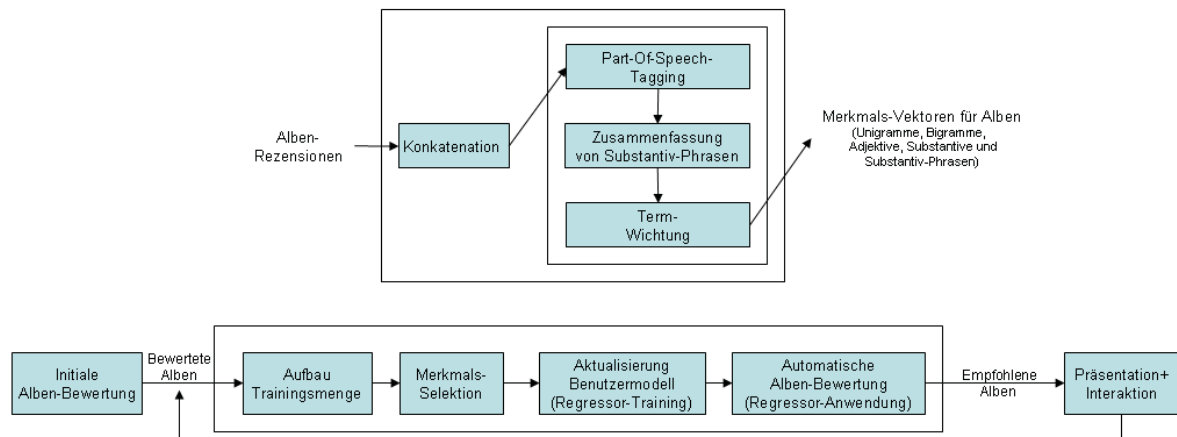


Abbildung 2.7: Maschinelles Lernen zur Revisions-basierten Alben-Empfehlung

nach relevanten Rezensionen, von denen auf die Relevanz der beschriebenen Stücke geschlossen wird.

Als Alternative zu diesem Ansatz wurde ein auf maschinellem Lernen basierendes Empfehlungssystem realisiert. Abbildung 2.7 zeigt die zentralen Prozesse zur Revisions-Modellierung und Empfehlungs-Generierung.

Wie bei Whitman wurde die alternative Verwendung unterschiedlicher Merkmale zur Reduktion von Volltexten auf gewichtete Term-Vektoren (Unigramme, Bigramme, Adjektive, Substantive und Substantiv-Phrasen) untersucht.

Die Menge der bewerteten Alben dient zum Aufbau eines Trainingskorpus, dessen hohe Dimensionalität eine Merkmalsselektion (mittels χ^2 - oder maximal-information-gain-Test) erforderlich macht. Diese so konzentrierten Daten können anschließend zum Lernen eines Regressors verwendet werden, der die Beziehung zwischen dem Term-Raum und der getroffenen Benutzer-Wertung modelliert. In [Yan08] wurden hierzu Support-Vector-Regressoren, k-Nearest-Neighbor-Verfahren und Naive-Bayes-Methoden (nach Quantifizierung der Benutzer-Bewertung in 10 Klassen) implementiert. Ein so gelernter Regressor beschreibt das Bewertungsverhalten des aktuellen Benutzers und kann somit zur Vorhersage der individuellen Relevanz anderer Texte (und damit Alben) herangezogen werden.

Zur vergleichenden Evaluation beider Verfahren wurden die generierten Empfehlungen auf eine Zugehörigkeit zur gleichen Amazon-Genre-Menge wie die positiv bewerteten Alben hin überprüft. Als zweite, spezifischere Quelle einer Grundwahrheit wurden direkte kollaborativ ermittelte Ähnlichkeits-Verbindungen von Amazon herangezogen. Details zur Definition der Qualitätsmaße finden sich in [Yan08].

Es zeigte sich sowohl in der Güte der gefundenen Empfehlungen als auch bzgl. der Performanz für deren Berechnung eine klare Überlegenheit des Information-Retrieval-basierten Ansatzes. Die Gültigkeit dieses Ergebnisses wurde ebenso wie die Messbarkeit einer sukzessiven Verbesserung des Empfehlungsqualität im Laufe des iterativen Feedback-Prozesses im Rahmen einer kleinen Nutzerstudie validiert [Yan08].

2.4.5 Diskussion

Textbasierte Ähnlichkeitsmaße verlagern die Aufgabe der für eine Ähnlichkeitsbestimmung notwendigen maschinellen Analyse und Interpretation von einer musikalischen in die natürlich-sprachliche Domäne, die auf eine längere Forschungsgeschichte zurückblicken kann. Diese Übertragung ist jedoch nicht verlustfrei: Eine lediglich durch weitere Schlüsselwörter wie „review“ oder „music“ eingeschränkte Anfrage an eine Internet-Suchmaschine mit mehrdeutigen Künstler-Namen wie „Texas“ oder „Cure“ resultiert in einer hohen Anzahl irrelevanter Informationen, die den Ähnlichkeitsbegriff verrauschen. Eine Beschränkung auf Datenquellen hoher Qualität (wie dezidierte Internet-Portale mit Alben-Rezensionen) kann dieses Problem zwar abschwächen, führt andererseits jedoch zu einer unerwünschten Reduktion der nutzbaren Datenmenge. Dies ist auch vor dem Hintergrund zu vermeiden, dass Rezensionen stets die subjektive Einordnung der beschriebenen Musik durch den Rezensenten darstellen. Ein ähnlicher Kompromiss zwischen Daten-Qualität und -Umfang muss auch in Bezug auf die Auswahl von relevanten Termen zur Modellierung eines Textes getroffen werden: ein festes Vokabular garantiert zwar die Konzentration auf musikalisch bedeutsame Begriffe, ist jedoch (gerade zur Beschreibung kreativer Prozesse wie der Musik-Produktion) zu unflexibel.

Whitman argumentiert, dass die Wahrnehmung und Einordnung von Musik nicht zeitinvariant und somit prinzipiell schlecht allein von inhaltsbasierten Maßen erfassbar sei. Textbasierte Ähnlichkeitsmaße, die auf aktuellen Internet-Crawling-Ergebnissen basieren, seien dagegen besser in der Lage, diese Dynamik der menschlichen Beurteilung von Musik zu berücksichtigen [WL02].

Die Basierung auf kontextuellen Metadaten, die über inhärente Informationen im Tonsignal hinausgehen, ermöglicht textuellen Ähnlichkeitsmaßen wie bereits den kollaborativen Ansätzen eine Berücksichtigung von kulturellen und „nicht-musikalischen“ Ähnlichkeitsdimensionen und den Zugriff auf semantisch höherwertige Konzepte wie Genre-Bezeichnungen oder Gefühls-Assoziationen, die ansonsten fehlerbehaftet automatisch ermittelt werden müssten.

Über die Verwendung als Basis eines Ähnlichkeitsmaßes hinaus haben die vorgestellten Verfahren, die ein gegebenes Album oder einen gegebenen Künstler mit textuellen Beschreibungen annotieren, eine große Relevanz für Music-Information-Retrieval-Ansätze, die dem *Query-By-Description*-Paradigma folgen, also zu einer natürlichsprachigen Anfrage passende Musik finden [KW07, TBTL07, WE04, WR02]. Wie in Abschnitt 1.2.3 ausgeführt, verlangt dies jedoch vom Benutzer die Formulierbarkeit seines Informationsbedürfnisses und resultiert letztlich in einem unnötigen durch Explorationsumgebungen vermeidbaren Medienbruch.

2.5 Zusammenfassende Bewertung

In Ergänzung zu den isolierten Diskussionen der verschiedenen Ansätze zur Ähnlichkeitsbestimmung in den einzelnen Unterkapiteln der vorangehenden Abschnitte, werden im folgenden die Charakteristika der unterschiedlichen Methoden zusammengefasst, gegenübergestellt und bewertet.

Hierzu stellt Tabelle 2.3 die Stärken und Schwächen inhaltsbasierter, Experten-basierter, kollaborativer und Text-basierter Verfahren hinsichtlich einer Reihe von Kriterien überblicksartig dar, die im folgenden näher beleuchtet werden.

Kriterium	Inhalts-basiert	Experten-basiert	Kollaborativ	Text-basiert
Anwendbarkeit und Abdeckung	++	--	–	-- / –
Skalierbarkeit	+	--	+	-- / O
(Kultureller) Kontext	--	+	++	O
Semantisches Niveau	--	++	+	O
Konzentration auf spezifische Einzelaspekte	++	–	–	–
Qualität (gemäß [LEB03])	–	++	+	O
Berücksichtigung dynamischer Meinungsbilder	--	–	+	– / +

Tabelle 2.3: Bewertende Gegenüberstellung verschiedener Ansätze zur Ähnlichkeitsbestimmung

Während eine Bewertung der ersten drei Methoden meist recht präzise möglich ist, ist die Beurteilung Text-basierter Ähnlichkeitsverfahren mitunter stark von der Quelle der analysierten Textrepräsentation abhängig, die von Experten erstellt (z.B. redaktionelle Alben-Rezensionen) oder auf kollaborativen Daten (z.B. Nutzer-Erfahrungsberichte) basieren kann. Falls eine dahingehende Differenzierung für die Bewertung eines Kriteriums notwendig ist, wird dies durch zwei durch einen Schrägstrich getrennte Angaben notiert.

Anwendbarkeit und Abdeckung

Das Kriterium der Anwendbarkeit untersucht die Frage, unter welchen Voraussetzungen die Ähnlichkeit zwischen zwei gegebenen Titeln berechenbar ist. Hieraus resultieren Unterschiede in der Abdeckung der Ähnlichkeitsmatrix.

In dieser Kategorie spielen inhaltsbasierte Verfahren ihre Überlegenheit aus, da sie bei Vorliegen repräsentativer Audio-Ausschnitte die Ähnlichkeit zwischen zwei Stücken algorithmisch quantifizieren.

Experten-basierte Ansätze dagegen erfordern die aufwändige manuelle Erstellung von Kategorisierungen oder die Formulierung expliziter Ähnlichkeits-Verbindungen zwischen Stücken. Daher ist zu erwarten, dass diese Ähnlichkeitsmaße stets nur für einen (recht kleinen) Teil einer betrachteten Musikkollektion verfügbar sind.

Kollaborative Verfahren basieren ebenfalls auf menschlichen Bewertungen oder der Beobachtung des Einkaufsverhaltens von Kunden. An die Stelle weniger, aber qualitativ hochwertiger Experten-Meinungen tritt hier die Aggregation einer Vielzahl von unkontrollierten Einzelmeinungen in der Hoffnung, dass sinnvolle Strukturen aus dem Rauschen emergieren. Daher setzen kollaborative Ähnlichkeitsberechnungen eine hinreichend umfangreiche Bewertungsbasis voraus. Dies ist insbesondere bei neu erscheinenden Musiktiteln problematisch, die sich somit einer kollaborativen Einordnung prinzipiell entziehen („new item problem“).

Text-basierte Verfahren weisen (je nach Datenquelle) die Problematik der Experten-basierten oder kollaborativen Ansätze auf.

Skalierbarkeit

Recht eng mit der Anwendbarkeit verknüpft ist das Problem der Skalierbarkeit der verschiedenen Ansätze, also die Frage, inwiefern die Verfahren auch bei größeren Datenbeständen eingesetzt werden können.

Erneut ergibt sich aus der automatisierbaren inhaltsbasierten Ähnlichkeitsermittlung eine positive Bewertung der Skalierbarkeit, wenn auch die Berechnung einzelner Maße mitunter komplex und zeitintensiv werden kann.

Die manuelle Annotation größerer Datenbestände durch Experten (wie dies z.B. im Rahmen des Music Genome Projects durchgeführt wird), stößt jedoch schnell an Grenzen.

Kollaborative Verfahren sind prinzipiell für die Verarbeitung großer Datenmengen ausgelegt (und erfordern diese wie im letzten Abschnitt erläutert sogar). Dennoch kann auch hier die Berechnungskomplexität ohne weitere Maßnahmen eine reaktive Ähnlichkeitsberechnung verhindern.

Erneut lassen sich Text-basierte Verfahren als Weiterverarbeitung von Experten- bzw. kollaborativen Daten interpretieren und somit die Bewertungen der entsprechenden Methoden übernehmen. Der Rechenaufwand für die sich anschließende Auswertung der kollaborativen Daten führt jedoch zu einer Herabsetzung der Beurteilung.

(Kultureller) Kontext

Inhaltsbasierte Verfahren beschränken sich naturgemäß auf eine Auswertung des klanglichen Materials von Musik. Die anderen drei Methoden zur Ähnlichkeitsberechnung, die auf menschlichen Äußerungen oder Handeln basieren, können dagegen auch Ähnlichkeitsbestimmende Faktoren berücksichtigen, die sich nur durch den (kulturellen) Kontext erklären lassen, in dem die Musik steht (Beispiele hierzu finden sich in Abschnitt 2.2). Besonders ausgeprägt tritt dieser Kontext in kollaborativen Maßen zu Tage. Text-basierte Verfahren haben zwar auch prinzipiell Zugriff auf diese Informationsquelle; die Ungenauigkeit der Extraktion dieses Wissens aus den vorliegenden natürlichsprachlichen Äußerungen führt jedoch zu einer geringeren Bewertung dieses Ansatzes.

Semantisches Niveau

Die Ermittlung höherwertiger semantischer Merkmale von Musik (wie z.B. Stimmung, „Tanzbarkeit“ o.ä.) mit signaltheoretischen Verfahren allein auf der Grundlage der Audiodaten gilt noch immer als nur unbefriedigend gelöstes Problem. Hier besitzen Experten-basierte Verfahren, die auf kompetenten Analysen der Musik beruhen, klare Vorteile. Auch kollaborative Ansätze berücksichtigen implizit höherwertige Merkmale, da sie auf der Beobachtung menschlichen Verhaltens beruhen. Erneut sind Text-basierte Verfahren aufgrund des Informationsverlustes als Resultat ihrer fehlerhafteten automatischen natürlichsprachigen Analyse schlechter zu bewerten.

Konzentration auf spezifische Einzelaspekte

Wie anhand der Klassifikation inhaltsbasierter Ähnlichkeitsmaße in Abschnitt 2.1 erkennbar, lassen sich klare musikalische Aspekte identifizieren, die von einigen Maßen isoliert betrachtet

werden und somit z.B. eine Beurteilung der Ähnlichkeit zwischen zwei Stücken hinsichtlich klanglicher Charakteristika von einer Ähnlichkeit hinsichtlich rhythmischer Merkmale abgrenzen. Dies ermöglicht die individuelle Zusammensetzung verschiedener Maße zu einem personalisierten Gesamtmaß (vergl. Abschnitt 2.6 und 4.2).

Diese Konzentration auf einzelne musikalische Aspekte ist bei den anderen Verfahren nicht gegeben. An die Stelle einer „feingranularen“ Analyse tritt ein vorgegebener allgemeingültiger Ähnlichkeitsbegriff.

Qualität (gemäß [LEB03])

Logan et al. bewerten die Qualität verschiedener Ähnlichkeitsmaße anhand ihrer Konsistenz mit Erwartungen einer Nutzer-Testgruppe. In diesen Untersuchungen wird eine klare Überlegenheit der bewusst von Experten definierten Maße beschrieben. Inhaltsbasierte Ansätze erreichen (zumindest nach dem aktuellen Forschungsstand) nicht dieses Niveau. Kollaborative Verfahren erfordern eine hinreichend umfangreiche Datenbasis, um qualitativ hochwertige Ähnlichkeitszusammenhänge sichtbar zu machen. Da aus natürlichsprachigen Texten *indirekt* auf musikalische Ähnlichkeit geschlossen werden muss, ist die Qualität Text-basierter Ähnlichkeitsfunktionen erneut niedriger zu bewerten.

Berücksichtigung dynamischer Meinungsbilder

Wie Whitman in [WL02] darlegt, ist die Beurteilung musikalischer Ähnlichkeit einer zeitlichen Veränderung unterworfen; eine Quantifizierung müsse auf diese Zeitvarianz reagieren und die Berechnung adaptieren.

Inhaltsbasierte Merkmale sind prinzipiell zeitinvariant, da sie rein auf den unveränderlichen Klangdaten der Stücke basieren. Die musikalischen Analysen und Klassifizierung von Musikstudierenden, die die Grundlage für Experten-basierte Maße darstellen, weisen in ähnlicher Weise einer eher geringen Änderungsfrequenz auf. Am flexibelsten reagieren kollaborative Verfahren auf ein sich wandelndes Ähnlichkeitsverständnis. Die Bewertung textbasierter Verfahren wird erneut durch die Beurteilung der zugrunde liegende Datenquelle bestimmt.

2.6 Kombination von Ähnlichkeitsmaßen

Die Stärken und Schwächen der vorgestellten Ähnlichkeitsmaße in unterschiedlichen Bereichen, die Subjektivität musikalischer Wahrnehmung und die Heterogenität persönlicher Musiksammlungen legen eine Nutzer-spezifische Kombination dieser Ähnlichkeitsmaße zur Modellierung eines personalisierten ganzheitlichen Ähnlichkeitsbegriffes nahe.

So untersuchen Baumann et al. den Einsatz eines individuell konfigurierbaren Maßes zum Aufbau eines mobilen Empfehlungssystems [BPV04]. Mit Hilfe eines Joysticks hat der Benutzer die Möglichkeit, die Gewichtung von drei Ähnlichkeitsaspekten zu kontrollieren: Neben einem inhaltsbasierten Maß, das durch Analyse des Frequenzspektrums die klanglichen Eigenschaften im Tonsignal berücksichtigt (vergl. Abschnitt 2.1.1) und einem kulturellen Ähnlichkeitsmaß auf der Grundlage von Alben-Rezensionen (vergl. Abschnitt 2.4.2) kann auf die textuelle Ähnlichkeit zugeordneter Liedtexte (vergl. Abschnitt 2.4.1) zurückgegriffen werden.



Abbildung 2.8: Drei verschiedene Gewichtungen von Musikähnlichkeits-Aspekten nach Baumann [BPV04]

Abbildung 2.8 zeigt beispielhaft die Merkmalskompositionen von drei verschiedenen identifizierten Benutzertypen. Baumann et al. nutzen ein einfaches lineares Modell zur Errechnung der Gesamt-Ähnlichkeit aus den Einzelähnlichkeiten:

$$sim_{Baumann}(s, t) = w_{sound} \cdot sim_{sound}(s, t) + w_{style} \cdot sim_{style}(s, t) + w_{lyrics} \cdot sim_{lyrics}(s, t),$$

wobei sich die Gewichte der Einzelähnlichkeitsmaße zu 1 addieren ($w_{sound} + w_{style} + w_{lyrics} = 1$).

In analoger Weise verwendet Jacobsen ein lineares Modell aus klangbasierter und rhythmischer Ähnlichkeit (basierend auf Beat Spectrums, vergl. Abschnitt 2.1.2) sowie dem Vergleich des inhaltsbasiert abgeleiteten Songaufbaus [Jac06]. Weitere Arbeiten, die eine durch den Benutzer bewusst gesteuerte Gewichtung verschiedener Ähnlichkeitsaspekte ermöglichen, sind [PG07, VP05].

Auf der Basis eines so kombinierten Gesamt-Ähnlichkeitsmaß kann anschließend eine personalisierte Visualisierung und Explorationsunterstützung der gegebenen Musiksammlung realisiert werden. Mit dieser Zielsetzung stellen Pampalk et al. einen alternativen Ansatz vor, der jedoch die interaktive Gewichtung verschiedener Ähnlichkeitsaspekte hinter die Visualisierungs-Generierung stellt [PDW03a]:

Ausgangspunkt sind die in Abschnitt 2.1.2 beschriebenen Periodicity Histograms zur Modellierung der rhythmischen Struktur eines Stückes sowie die in Abschnitt 2.1.1 vorgestellten Spectrum Histograms, die klangliche Eigenschaften des Tonsignals modellieren. Diese beiden Ähnlichkeitsmaße werden ergänzt um eine Metadaten-basierte Gruppierung, etwa eine Klassifikation nach Genres. Für jeden dieser Aspekte wird eine Self-Organizing Map generiert, die die Titel der Kollektion Topographie-erhaltend in einen zweidimensionalen Raum einbettet; ähnliche Stücke werden so auf dicht beieinander liegenden Zellen der Karte positioniert.

Zur Visualisierung von Kombinationen der Ähnlichkeitsmaße ergänzen Pampalk et al. diese drei SOMs zu einem Stapel sog. *aligned SOMs*, die verschiedene Zwischenstufen der Gewichtung der Einzelaspekte visualisieren. Die Erstellung dieser Merkmalskarten wird durch eine Ausdehnung des Nachbarschaftsbegriffes im Anpassungsprozess der SOM-Generierung realisiert, der nun auch Zellen von Karten ähnlicher Merkmalskombinationen im Stapel umfasst, und führt damit zu einer Angleichung der Titel-Positionierungen zwischen SOMs ähnlicher Aspekt-Gewichtung.

Den Vorteil einer interaktiven Exploration, bei der Auswirkungen einer Neugewichtung der verschiedenen Ähnlichkeitsmaße direkt sichtbar werden, erkauft man sich durch eine ressourcenintensive Vorab-Berechnung einer begrenzten Anzahl vordefinierter Merkmalskombinationen.

Obige Ansätze erfordern die bewusste Konfiguration von Merkmalsgewichten durch den Benutzer und setzen daher ein Wissen des Anwenders um die Semantik der Einzel-Aspekte voraus.

In dieser Arbeit tritt dagegen an die Stelle einer bewussten Parametrisierung der Ähnlichkeitsmaß-Kombination die automatische Anpassung des Systems an das Ähnlichkeitsverständnis des Nutzers. Hierzu wird aus der Beobachtung des Anwender-Verhaltens in der Explorationsumgebung, das sich in der Repositionierung von Stücken oder der Veränderung von Grenzziehungen zwischen Song-Gruppen zeigt, eine Neugewichtung erlernt. Kapitel 4 beschreibt im Detail diesen Adaptionsprozess.

2.7 Inhaltsbasierte Simulation kollaborativer Ähnlichkeit

Zahlreiche Studien belegen die höhere Qualität kollaborativer Ähnlichkeitsmaße gegenüber inhaltsbasierten Ansätzen [SK05, MS08a, BLEW04]. Wie in Abschnitt 2.3.4 diskutiert leiden kollaborative Verfahren jedoch unter dem *new item problem*: Neue Objekte im System, zu denen noch keine verwertbaren Nutzungskontexte vorliegen, sind prinzipiell aus der Menge der generierbaren Empfehlungen ausgeschlossen. Dies ist insbesondere daher relevant, da die Einordnung von Neuerscheinungen zu den Interessensprofilen der Kunden eine hohe ökonomische Bedeutung darstellt. Im Gegensatz dazu sind inhaltsbasierte Maße bei Vorliegen des Tonsignals stets automatisiert berechnbar und können direkt in Empfehlungssystemen Anwendung finden.

Stenzel und Kamps stellen einen Ansatz vor, der zum Ziel hat, die Qualität kollaborativer Maße durch inhaltsbasierte zu simulieren und somit die Vorteile beider Strategien zu kombinieren [SK05]. Hierzu induzieren sie mit Hilfe Maschinelles Lernverfahren ein Modell, das versucht, aufgrund von inhaltsbasierten Merkmalen die zu erwartenden Reaktion der Community vorherzusagen.

Grundlage ist ein Trainings- und Evaluationskorpus aus insgesamt 880 995 Playlists mit 36 916 Stücken, aus dem sich gemäß Abschnitt 2.3.2 ein kollaborativer Ähnlichkeitsbegriff ableiten lässt. Diesen Daten werden 30 mit Hilfe des Marsyas-Toolkits [TC00] ermittelte Klang- und Rhythmus-basierte Merkmale gegenübergestellt, die sich in [TC02] als nützlich zur Genre-Klassifikation erwiesen haben.

Zur Vereinfachung des späteren Lernverfahrens und zur Vermeidung einer Überanpassung wird die Menge der Playlists zunächst mit Hilfe eines unüberwachten k-Means-Clustering-Verfahrens in 200 disjunkte Gruppen partitioniert, von denen man annimmt, dass sie einen bestimmten Aspekt musikalischer Präferenzen modellieren.

Anschließend wird für jedes Stück in der Trainingsmenge und jedes Cluster ermittelt, ob es hinreichend häufig in einer Playlist der Gruppe auftritt. Dies resultiert letztlich in einer Repräsentation der kollaborativen Trainingsdaten eines Titels als 200-dimensionaler Vektor binärer Cluster-Zugehörigkeiten.

Stenzel und Kamps verwenden Support-Vector-Machines mit radialem Kernel zur Modellierung der Abhängigkeit zwischen den inhaltsbasierten Marsyas-Merkmalen und den Komponenten dieses Vektors. In ihrer Evaluation konnten sie zeigen, dass ein Empfehlungssystem auf der Basis dieser gelernten Kombinationen von Tonsignal-Merkmalen eine durchschnittlich nahezu doppelt so hohe Präzision wie ein rein inhaltsbasiertes Verfahren aufweist, wenn man obige Playlist-Co-Occurrence als Referenzähnlichkeit annimmt.

In ähnlicher Weise nutzen Whitman und Ellis inhaltsbasierte Maße zur Simulation eines kollaborativen Ähnlichkeitsbegriffes auf Künstler-Ebene [WE04], der jedoch nicht auf dem gemeinsamen Auftreten von Songs in Playlists fußt, sondern auf einer Analyse beschreibender textueller Metadaten (vergl. Abschnitt 2.4) basiert. An die Stelle einer schwierig zu interpretierenden Cluster-Zugehörigkeit tritt daher die Vorhersage eines Term-Vektors, der die Relevanz vorgegebener musikalischer Begriffe zur Charakterisierung eines Stückes bestimmt und als Vorstufe für ein Query-By-Description-System betrachtet werden kann.

Die Ermittlung dieser Term-Relevanzen wird auf der Grundlage einer inhaltsbasierten Merkmalsmenge, die im Wesentlichen spektrale Eigenschaften und deren zeitliche Veränderungen im Stückverlauf erfasst, mit Hilfe einer *Regularized Least-Squares Classification* [Vap98] realisiert, die als eine effizientere Alternative zu den von Stenzel und Kamps genutzten Support-Vector-Maschinen bei nur geringem Qualitätsverlust [Rif02] angesehen werden kann und zudem eine performante Aktualisierung des gelernten Modells erlaubt.

Wie in Abschnitt 2.4 beschrieben, lässt sich die Ähnlichkeit von zwei Stücken schließlich durch den Cosinus des Winkels zwischen den Termrelevanz-Vektoren quantifizieren.

Der Ansatz von Turnbull et al. unterscheidet sich im Wesentlichen in der Ersetzung der Regularized Least-Squares Classification durch ein stochastisches Modell, das die Wahrscheinlichkeitsdichte eines Term-Vektors in Abhängigkeit von einer inhaltsbasierten Merkmalsausprägung beschreibt [TBTL08]:

Hierzu wird die bedingte Wahrscheinlichkeit $P(\vec{x}|t_i)$ eines Features-Wertes $\vec{x}$ bei beobachtetem Einzel-Term t_i als Gauß'sche-Mischverteilung aufgefasst, deren Parameter durch eine Variante des EM-Algorithmus geschätzt werden.

Auf der Basis des Satzes von Bayes (und einiger Unabhängigkeitsannahmen) lässt sich hieraus die Wahrscheinlichkeit $P(t_i|\mathbf{x})$ des Terms t_i bei beobachteter Merkmalsmenge $\mathbf{x} = \{\vec{x}_1, \dots, \vec{x}_R\}$ wie folgt berechnen:

$$P(t_i|\mathbf{x}) = \frac{(\prod_{r=1}^R P(\vec{x}_r|t_i))^{\frac{1}{R}}}{\sum_{v=1}^V (\prod_{r=1}^R P(\vec{x}_r|t_v))^{\frac{1}{R}}},$$

wobei V die Größe des betrachteten Begriffs-Vokabulars beschreibt.

Aus diesen Einzelwahrscheinlichkeiten leiten Turnbull et al. schließlich die gesuchte Verbundwahrscheinlichkeit für einen Termvektor bei gegebener inhaltsbasierter Charakterisierung ab. Die Kullback-Leibler-Divergenz erlaubt schließlich eine Quantifizierung der Ähnlichkeit zweier Stücke durch einen Vergleich ihrer so ermittelten Wahrscheinlichkeitsverteilungen.

Die drei bislang diskutierten Ansätze gehen stets von einer Repräsentation der inhaltsbasierten Charakteristika in einem Euklidischen Vektorraum aus. Die Überführung dieser Merkmalsausprägungen in Vektoren eines semantisch hochwertigeren Raums (Playlist-Cluster-Zugehörigkeit bzw. Term-Gewichtungen), in dem schließlich die Ähnlichkeitsberechnung stattfindet, kann somit als eine Art Feature-Transformation angesehen werden. Dieser Ansatz schließt die Nutzung von komplexeren psychoakustisch motivierten Ähnlichkeitsmaßen wie z.B. den Klang-basierten Maße von Logan/Salomon und Aucouturier/Pachet (vergl. Abschnitt 2.1) aus, die lediglich paarweise Distanzen zwischen Stücken und keine Koordinaten eines niedrig-dimensionalen Euklidischen Raumes ermitteln.

In einer Diplomarbeit, die im Kontext dieser Dissertation entstand, wurde daher untersucht, inwiefern ein kollaboratives Ähnlichkeitsmaß aus der Kombination von inhaltsbasierten Ähnlichkeitsmatrizen gelernt werden kann [Kel07]. Hier wird die Ähnlichkeitsberechnung somit

bereits in der akustischen Domäne durchgeführt: Anstatt eine Feature-Transformation zu lernen, wird ein Modell zur Kombination der akustischen Ähnlichkeiten induziert.

Zur Aquisierung einer hinreichend umfangreichen Trainings- und Testdatenmenge kollaborativer Ähnlichkeit wurden die über WebServices zugreifbaren Wiedergabelisten von Last.fm genutzt. Zunächst wurden rund 20 Millionen über Freundschafts- und Nachbarschafts-Beziehungen verlinkte Benutzernamen ermittelt. Anschließend wurden die 50 meist gehörten Titel von 1,3 Millionen aktiven Benutzern dieses Kollektivs ausgelesen. Die 65 Millionen Einträge dieser Anwender-Playlists führen fast 100 000 verschiedene Stücke auf. Um eine Evaluation verschiedener Lernverfahren zu ermöglichen, beschränkt sich Keller auf die Titel, die mindestens 3000 mal auftreten. Dies reduziert die Datenmenge auf 1065 Lieder, die eine durchschnittliche Co-Occurrence von 64 aufweisen und somit eine brauchbare Quelle für ein kollaboratives Ähnlichkeitsmaß darstellen.

Wie in Abschnitt 2.3.2 bereits beschrieben, normiert Keller zur Quantifizierung der Ähnlichkeit zwischen den Stücken s und t deren Co-Occurrence $coc(s, t)$ durch den Mittelwert der Anzahl Playlists, auf denen sie erscheinen ($occ(s)$ bzw. $occ(t)$):

$$sim_{Keller}(s, t) = \frac{coc(s, t)}{\frac{1}{2}(occ(s) + occ(t))}$$

Der paarweise Vergleich von Titeln führt zu einer symmetrischen Ähnlichkeitsmatrix mit 1065^2 Einträgen, deren Elemente auf der Haupt-Diagonale 1 sind.

Diesem kollaborativen Maß wird eine Reihe von inhaltsbasierten Ähnlichkeitsmaßen gegenübergestellt, die unterschiedliche akustische Teilaspekte eines Musiktitels abdecken: Keller verwendet insgesamt 16 populäre Ansätze, die von einfachen Klang-basierten Merkmalen wie dem Spectral Centroid (vergl. Abschnitt 2.1.1) bis zu komplexeren Größen wie Beat-Histogrammen (vergl. Abschnitt 2.1.2) und psychoakustischen Maßen nach Logan/Salomon sowie Aucouturier/Pachet (vergl. Abschnitt 2.1) reichen.

Schließlich untersucht Keller, wie genau sich durch Kombination dieser 16 inhaltsbasierten Ähnlichkeitsmaße die kollaborative Referenz approximieren lässt. Diese Aufgabe kann man als Regressionsproblem formulieren: Gesucht ist eine Verknüpfung der akustischen Ähnlichkeiten, so dass die durchschnittliche quadratische Abweichung von sim_{Keller} zur kollaborativen Referenz für jeden paarweisen Vergleich von Stücken der Trainingsmenge minimal wird.

Keller betrachtet in seiner Arbeit eine Reihe von linearen Regressoren und stellt diese verschiedenen Support-Vector-Regressoren gegenüber. Abbildung 2.9 zeigt die Wurzel des quadrierten Approximationsfehler auf einer unabhängigen Testmenge bei Verwendung einer zufälligen Kombination sowie eines linearen und eines Support-Vector-Regressors in Abhängigkeit vom Umfang der Trainingsdaten. Wie sich erkennen lässt, sind beide Regressoren signifikant besser als der Zufall; der Unterschied zwischen beiden Ansätzen ist jedoch gering.

Ähnliche Schlussfolgerungen lassen sich aus Abbildung 2.10 ziehen, die die Korrelation der vorhergesagten Ähnlichkeit mit dem tatsächlichen kollaborativen Testwert zeigt. Der erreichbare Korrelationskoeffizient von niedrigen 0,16 zeigt zwar die Grenzen der Simulierbarkeit des Community-basierten Maßes auf, ist jedoch signifikant höher als der maximale Korrelationskoeffizient von 0,09, der mit einem einzigen inhaltsbasierten Ähnlichkeitsmaß auf der verwendeten Testdatenmenge erreichbar ist. Dieses positive Ergebnis wird durch eine kleine Benutzerstudie mit 7 Teilnehmern bestätigt, in der zu einem gegebenen Stück eine vom

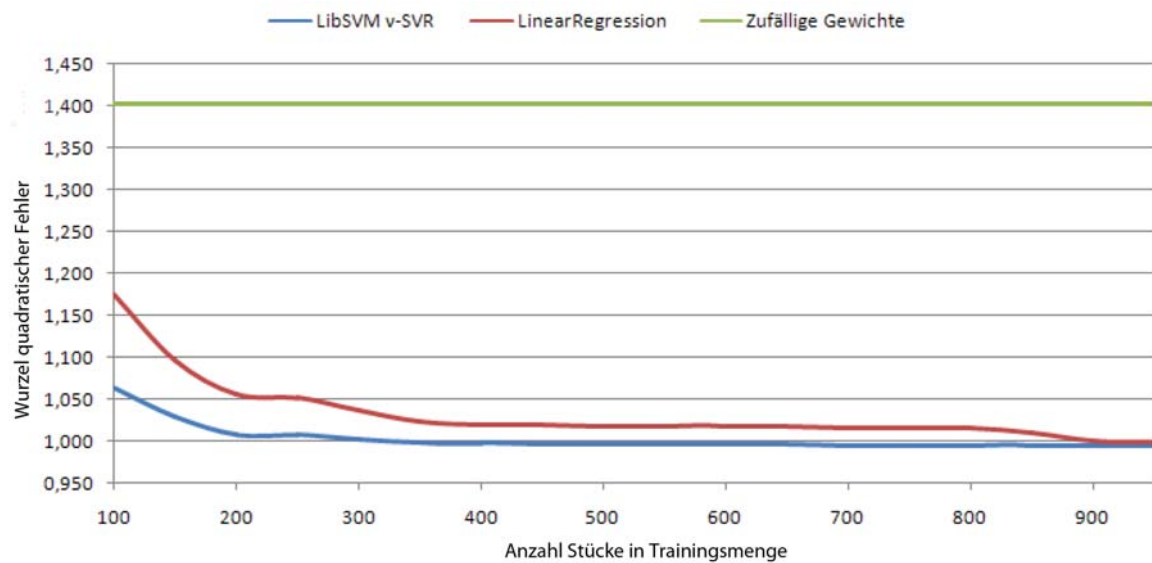


Abbildung 2.9: Wurzel des quadrierten Fehlers für zufällige Support-Vector-Regressor, linearen Regressor und zufällige Kombination in Abhängigkeit von der Größe der Trainingsmenge [Kel07]

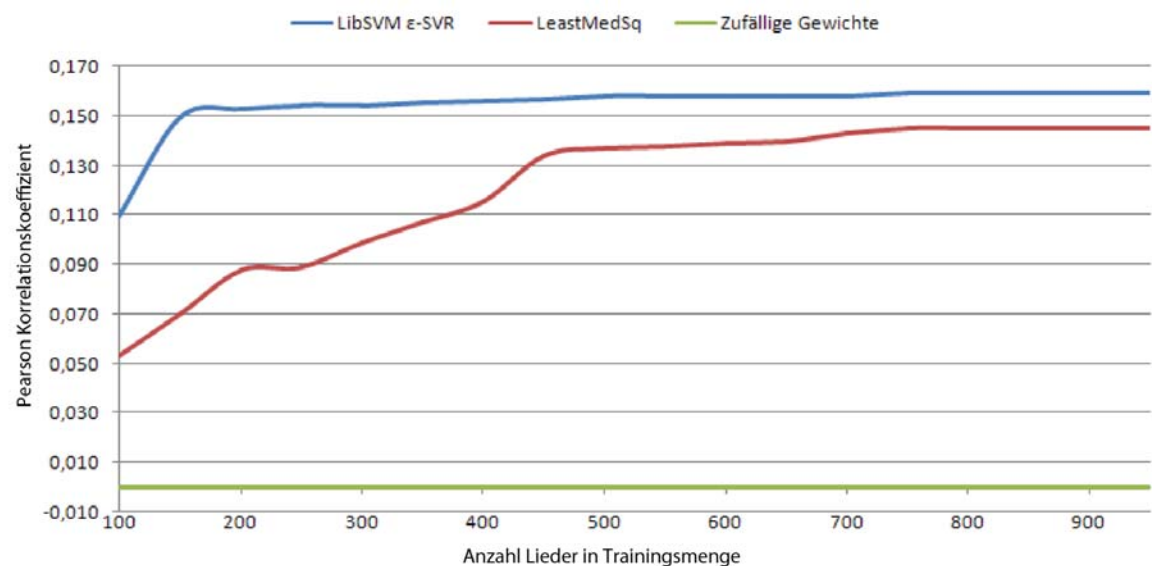


Abbildung 2.10: Korrelationskoeffizient nach Pearson zwischen kollaborativer Ähnlichkeit und vorhergesagter inhaltsbasierter Ähnlichkeit mittels Support-Vector-Regressor, linearem Regressor und zufälliger Kombination in Abhängigkeit von der Größe der Trainingsmenge [Kel07]

System nach Ähnlichkeit sortierte Titelliste mit der Einstufung des Probanden verglichen wurde.

Kellers feste Referenzähnlichkeit ist ein kollaboratives Maß, das auf der Beobachtung des Hörverhaltens vieler Benutzer basiert und somit eine Art Durchschnitts-Meinung repräsentiert und die Subjektivität des Ähnlichkeitsbegriffes unberücksichtigt lässt. In dieser Dissertation wird daher eine personalisierte Kombination von Ähnlichkeitsmaßen zur Modellierung des Organisationsprinzips des Anwenders verwendet.

2.8 Ähnlichkeitsmaße im soniXplorer

Musikalische Ähnlichkeit ist ein facettenreicher Begriff. Obige Darstellungen beschreiben verschiedene Möglichkeiten zur Quantifizierung unterschiedlicher Teilaspekte. Die Subjektivität eines Ähnlichkeitsverständnisses verlangt die personalisierte Gewichtung dieser Facetten.

Zum Aufbau eines Explorations-Prototypen, der eine Benutzer-adaptive Kombination verschiedener Aspekte ermöglicht, wurde eine Auswahl an heterogenen Ähnlichkeitsmaßen getroffen, die in die Realisierung des soniXplorers Einzug gefunden haben. Als Kriterium diente neben der Diversität der Ähnlichkeitsmaße auch die Verfügbarkeit einer validen Implementierung.

Eine offene Software-Architektur mit klar definierten Schnittstellen ermöglicht darüber hinaus die einfache Integration weiterer Ähnlichkeitsmaße, sofern diese nur für zwei gegebene Stücke einen Distanz-Wert berechnen. Somit können die im Rahmen dieser Arbeit realisierten Ähnlichkeitsmaße als exemplarische Implementierungen angesehen werden.

2.8.1 Inhaltsbasierte Merkmale

Klangliche Ähnlichkeit, die aus dem Audio-Signal abgeleitet werden kann, wird mit Hilfe der Maße nach Logan/Salomon und Aucouturier/Pachet ermittelt (vergl. Abschnitt 2.1.1).

Aufgrund ihrer aufwändigen psychoakustischen Modellierung wurden der Vergleich von *Fluctuation Patterns* (2.1.2), *Periodicity Histograms* (2.1.2) und *Spectrum Histograms* (2.1.1) nach Pampalk et al. integriert.

Mörchen et al. untersuchen in [MUTL06] insgesamt rund 66.000 inhaltsbasierte Merkmale, die sich aus der Kombination von ca. 400 Fenster-basierten Features und verschiedenen Aggregationsverfahren ergeben. Qualitätskriterium für eine überwachte Auswahl performanter Größen aus dieser riesigen Merkmalsmenge ist ein sog. *Separation Score*, der die Trennbarkeit von Stücken aus 5 verschiedenen Genre-Klassen in der Trainingsdatenmenge mit Hilfe des betrachteten Features quantifiziert. Ein in den Selektions-Prozess integrierter Korrelationsfilter garantiert ferner eine geringe Redundanz der Merkmalsmenge.

Die 20 so ermittelten besten inhaltsbasierten Features werden von Mörchen et al. zum Aufbau einer SOM-basierten Visualisierung von Musikkollektionen, dem sog. *MusicMiner*, verwendet ([MUNS05], vergl. Abschnitt 3.1.4). Da sich die Anwendungskontexte dieses Projektes und des soniXplorers ähneln und die ermittelte Merkmalsmenge als deutlich performanter im Vergleich zu anderen populären Features beschrieben wird [MUTL06], wurde der Euklidische Abstand zwischen den 20-dimensionalen Vektoren zweier Stücke als MusicMiner-Distanz in den soniXplorer-Katalog aufgenommen.

Eine letzte Gruppe integrierter inhaltsbasierter Merkmale basiert auf Berechnungen der Signalverarbeitungs-Bibliothek *mirToolbox* [LT07]. Hierzu zählen die in Abschnitt 2.1.1 diskutierten klanglichen Charakteristika *Zero crossing rate*, *Irregularity*, *Spectral centroid*, *Brightness*, *Spectral rolloff* bei 95% und 85%, *Spectral spread*, *Spectral flux*, *Spectral flatness* und *Key clarity*.

Darüber hinaus wurden Features, die den rhythmischen Gehalt eines Stückes beschreiben, aufgenommen. Neben dem RMS-Wert und der *Low energy rate* (vergl. Abschnitt 2.1.2) umfasst dies eine Tempo-Detektion, die durch eine Analyse von Periodizitäten ermittelter Anfangszeitpunkte („onsets“, vergl. Abschnitt 2.1.2) von Tönen im Signal realisiert wird [Lar08].

Zur Zusammenfassung dieser Fenster-basierten Größen werden die Feature-Werte durch Mittelwert-Bildung und Berechnung der Standardabweichung zu Stück-weiten Merkmalen aggregiert, deren Ausprägungen mit Hilfe der Euklidischen Distanz verglichen werden.

2.8.2 Metadaten-basierte Merkmale

Neben diesen inhaltsbasierten Charakteristika betrachtet *soniXplorer* bibliographische Metadaten über die zu vergleichenden Stücke.

Zum einen wird der Zeitraum zwischen den Veröffentlichungszeitpunkten beider Titel (bzw. der Alben, auf denen sie erscheinen) berechnet. Zur Gruppierung von Stücken des gleichen Künstlers wird ferner die Künstler-Bezeichnung betrachtet. Wie in [WL02] und [PL01] beschrieben sind leicht unterschiedliche Schreibweisen des gleichen Namens nicht selten. Auch Keller beschreibt eine signifikante Erhöhung der Playlist-Cooccurrence nach Elimination von Tippfehlern in Künstler- und Titel-Bezeichnungen [Kel07]. Anstatt eines einfachen binären Vergleiches wird daher die Edit-Distanz (mit identischen Kosten für Buchstaben-Änderung, -Einfügung und -Löschung) berechnet.

Als Quelle dieser Metadaten dienen zunächst die Felder der ID3-Annotation des MP3s. Falls dort keine gültigen Informationen gefunden werden können, wird ein Audio-Fingerabdruck des Stückes generiert und der Titel mit Hilfe des MusicDNS Webservice von Music IP¹¹ identifiziert. Fehlende oder offensichtlich inkorrekte ID3-Metadaten können auf der Grundlage der ermittelten eindeutigen Stück-ID anschließend anhand der Kataloge von MusicBrainz¹² oder Last.fm automatisch ergänzt bzw. korrigiert werden.

2.8.3 Kollaborative Merkmale

Die Qualität kollaborativer Ähnlichkeitsmaße hängt vom Zugriff auf eine hinreichend große Datenmenge über das Hörverhalten von Musiknutzern ab. Die in *soniXplorer* integrierten Distanzberechnungen nutzen daher die Webservices von Last.fm und Art of the Mix, die umfangreiche Informationen über ihren Kundenstamm zur Verfügung stellen.

Die folgenden Abschnitte beschreiben die implementierten Distanzmaße im Detail.

Direkte Ähnlichkeitsverknüpfung

Auf der Basis der Abfolge von mit registrierten Medienplayern seiner Benutzer wiedergegebenen Titeln und dem expliziten Feedback von Anwendern auf die Zusammenstellung von

¹¹<http://www.musicip.com/dns/index.jsp>

¹²<http://musicbrainz.org/>

personalisierten Radio-Streams generiert Last.fm direkte Ähnlichkeitsverknüpfungen zwischen Künstlern und einzelnen Titeln.

Für ein gegebenes Stück lässt sich eine Menge in diesem Sinne ähnlicher Titel mit Hilfe des `Track.getSimilar`-Webservices ermitteln. Last.fm quantifiziert die Ähnlichkeit der zurückgegebenen Songs durch einen *match*-Score, der dem ähnlichsten Titel einen Wert von 100 zuordnet. Aufgrund des beschränkten Umfangs der Menge kann bereits das Auftreten eines Stückes t in der Liste ähnlicher Titel zu s als Indiz einer starken Ähnlichkeit zwischen s und t betrachtet werden. Anstatt einer linearen Skalierung des match-Wertes werden daher mit Hilfe einer Exponential-Funktion bereits kleine match-Scores auf überproportional kleine Distanzen abgebildet. Ferner wird zur Sicherstellung der Symmetrie das arithmetische Mittel der gegenseitigen Ähnlichkeiten herangezogen. Somit errechnet sich die Distanz aufgrund einer direkten Last.fm-Ähnlichkeitsverknüpfung zwischen den Stücken s und t wie folgt:

$$d_{DV}(s, t) = \frac{1}{2} \cdot (e^{-c_{DV} \cdot \frac{ms_s(t)}{100}} + e^{-c_{DV} \cdot \frac{ms_t(s)}{100}}),$$

wobei $ms_s(t)$ den match-Score des Stückes t in der zurückgegebenen Ähnlichkeitsliste zu s darstellt bzw. 0 ist, falls t dort nicht auftritt.

Abschließend wird d_{DV} auf einen Wertebereich zwischen 0 und 1 normiert:

$$d_{DV_{norm}}(s, t) = \frac{d_{DV}(s, t) - e^{-c_{DV}}}{1 - e^{-c_{DV}}}$$

Die Konstante c_{DV} quantifiziert die Stärke der Nicht-Linearität der Transformation und wurde empirisch auf $c_{DV} = 5$ gesetzt.

Die Betrachtung direkter Ähnlichkeitsverbindungen zwischen Titeln stellt ein sehr spezifisches Maß dar, für dessen sinnvolle Berechnung selbst die große Last.fm-Datenbasis mitunter nicht hinreichend ist. Daher kombiniert `soniXplorer` die Stück-basierte Distanz linear im Verhältnis 2:1 mit einer analog berechneten Distanz auf Künstler-Ebene, die eine validere aber gröbere Schätzung ermöglicht.

Top Tags

Wie in Abschnitt 2.3.3 beschrieben ermöglicht Last.fm seinen Nutzern die Annotation von Stücken und Künstlern mit frei wählbaren Schlagwörtern. Eine Liste der zur Beschreibung eines Objektes am häufigsten verwendeten Tags kann durch einen Webservice abgefragt werden und wird daher als Quelle für ein weiteres kollaboratives Distanzmaß genutzt.

Zum Aufbau von Term-Vektoren, die mit Hilfe des klassischen Cosinus-Maßes verglichen werden können, ist die Berechnung der Gewichtung der assoziierten Schlagwörter notwendig, die auf Term- und Dokumentfrequenz aufbaut. Leider gibt Last.fm auf einen Webservice-Aufruf zwar eine nach ihrer Verwendungs-Häufigkeit sortierte Tag-Liste zurück, deren `count`-Attribut quantifiziert jedoch die Titel-unspezifische allgemeine Popularität des Schlagwortes und ist daher zur Berechnung der Termfrequenz unbrauchbar.

Fasst man die einem Titel zugeordneten Schlagwörter als Terme eines natürlichsprachigen Textes auf, so lässt sich das Zipf'sche Gesetz anwenden, das die Schätzung der Frequenz eines Wortes aus seinem Rang in einer nach der Auftretenshäufigkeit sortierten Liste ermöglicht

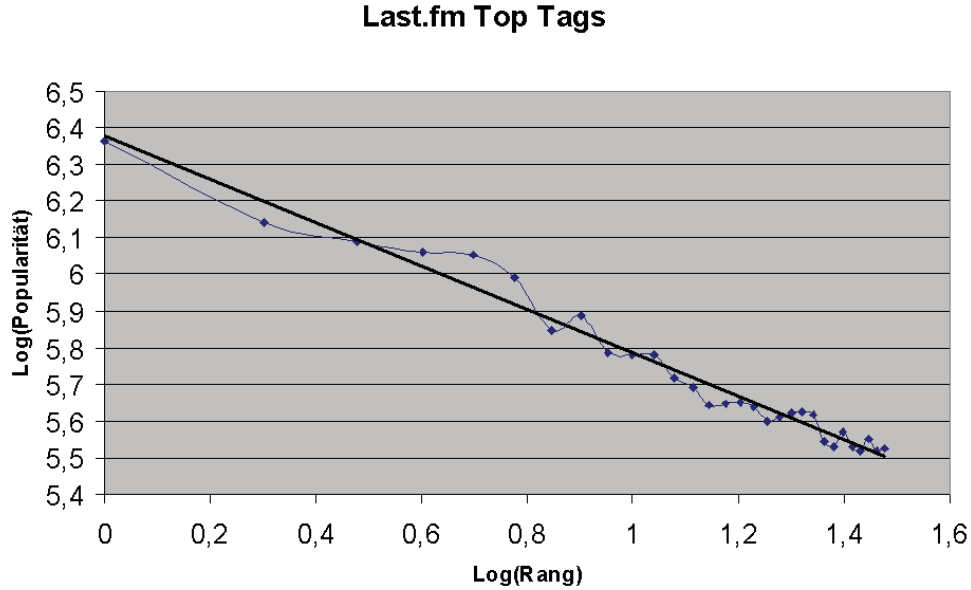


Abbildung 2.11: Gegenüberstellung der Popularität der ersten 30 am häufigsten bei Last.fm genutzten Schlagwörter und ihrem Rang

[MS99]. Inwiefern das Gesetz auch im Kontext von Last.fm-Tags anwendbar ist, lässt sich aufgrund des fehlenden Zugriffs auf die tatsächlichen Schlagwort-Häufigkeiten nicht endgültig beurteilen.

Ein anderer Last.fm-Webservice erlaubt jedoch die Zuordnung der Gesamt-Popularität eines Tags zu seinem Rang in der Liste der insgesamt am häufigsten verwendeten Schlagwörter. Die Visualisierung der Gegenüberstellung der ersten 30 Tags in einem doppelt-logarithmischen Diagramm (vergl. Abbildung 2.11) zeigt einen näherungsweisen linearen Verlauf und rechtfertigt damit eine Approximation der Tagfrequenz durch die Zipf'sche Dichte wie folgt:

$$tf_{tag,s} = 1/(\text{rank}_{tag,s} \cdot \ln(1,78 \cdot N_s)),$$

wobei $\text{rank}_{tag,s}$ den Rang des Schlagwortes tag in der nach Häufigkeit sortierten Liste der N_s Tags zum Stück s beschreibt.

Zur Gewichtung eines Terms gemäß dem $tf \cdot idf$ -Schema ist ferner die Dokumentfrequenz, d.h. der Anteil der Dokumente, die einen gegebenen Term enthalten, erforderlich. Last.fm offeriert jedoch keinen Webservice, mit dem sich die Anzahl der mit einem Tag annotierten Stücke ermitteln ließe. Zur Approximation wird daher die im `count`-Attribut zurückgegebene Gesamt-Popularität eines Schlagwortes t herangezogen, die die Häufigkeit der Verwendung von t im gesamten Last.fm-System quantifiziert. Trotz ihrer Ungenauigkeit, die sich aus der fehlenden Aggregation dieser Annotations-Zählungen zu binären Entscheidungen auf Stück-Ebene ergibt, kann diese Größe zur groben Schätzung der Spezifität eines Schlagwortes verwendet werden.

Hierzu ist die Popularität eines Tags in Relation zur Anzahl insgesamt durchgeführter Annotationen zu setzen. Da auch diese Größe nicht direkt von Last.fm ermittelbar ist, wird sie anhand der Popularität des prominentesten Schlagwortes "rock" abgeschätzt.

Somit errechnet sich die Stück-Frequenz trf_{tag} eines Tags tag wie folgt:

$$trf_{tag} = \frac{\text{pop}_{tag}}{c_{tt} \cdot \text{pop}_{\text{"rock"}}}.$$

Erste Experimente zeigten, dass die Wahl der Konstante c_{tt} die Ähnlichkeitsberechnung nicht maßgeblich beeinflusst. Mit der Annahme, dass jede zehnte Annotation eine Verschlagwortung mit "rock" ist, wurde $c_{tt} = 10$ gesetzt.

Tag-Frequenz $tf_{tag,s}$ und Stück-Frequenz trf_{tag} können damit nach klassischem $tf \cdot idf$ -Schema zur Berechnung von Schlagwort-Gewichten genutzt werden. Die Ähnlichkeit zwischen so aufgebauten ausgerichteten Tag-Vektoren lässt sich schließlich mit Hilfe des Cosinus-Maßes quantifizieren und in eine Distanz umwandeln.

Mit der gleichen Begründung wie im vorangehenden Abschnitt werden Tag-Distanz auf Stück- und Künstler-Ebene linear im Verhältnis 2:1 zu einem Gesamt-Abstand kombiniert.

Gemeinsames Auftreten auf Playlists

Das letzte kollaborative Distanzmaß, das in soniXplorer integriert wurde, basiert auf der Analyse von publizierten Wiedergabelisten des Portals Art of the Mix. Wie in Abschnitt 2.3.2 beschrieben trifft man die Annahme, dass zwei Stücke, die gemeinsam auf einer Playlist auftreten, den gleichen Musikgeschmack bedienen und somit ähnlich zueinander sind.

Das dort diskutierte Überdeckungsmaß $sim_{Luebbers}(s, t)$ kann aufgrund seines beschränkten Wertebereiches einfach zu einem Distanzmaß transformiert werden:

$$d_{AotM}(s, t) = 1 - sim_{Luebbers}(s, t)$$

Wie bei den anderen kollaborativen Maßen wird auch d_{AotM} mit einer analogen Distanz auf Künstler-Ebene kombiniert.

2.8.4 Fehlende Werte

Die Berechnung der paarweisen Abstände zwischen allen Stücken der zu visualisierenden Musikkollektion gemäß den oben vorgestellten Maßen führt zum Aufbau eines Stapels von Distanzmatrizen, die geeignet zu einem Gesamt-Abstandsbegriff kombiniert werden müssen.

Die nur geringfügig schlechtere Performanz eines linearen Modells zur Distanz-Kombination im Vergleich zu aufwändigeren Support-Vector-Machines, die in [Kel07] gezeigt wurde (vergl. Abschnitt 2.7), begründet die Berechnung der Distanz-Kombination als gewichtete Summe der Einzel-Matrizen. Die initiale Parameterisierung dieses Modells basiert auf der Gewichtung inhaltsbasierter Merkmale, die in [Kel07] zur Simulation eines kollaborativen Maßes ermittelt wurde. Kapitel 4 beschreibt die schrittweise Adaption dieser Distanzkombination an das Ähnlichkeitsverständnis des Benutzers im Verlauf des Explorationsprozesses.

Voraussetzung für eine Summation ist das Vorliegen aller Teil-Distanzen. Während man von der Berechenbarkeit der inhaltsbasierten Distanzen für alle Stücke ausgehen kann, ist dies für die Metadaten-basierten und kollaborativen Merkmale nicht notwendigerweise gegeben, da ein Titel möglicherweise nicht identifiziert werden kann oder die ermittelbaren kollaborativen Informationen nicht hinreichend umfangreich sind.

Eine Möglichkeit zur Lösung dieses Problems besteht in der Ersetzung fehlender Werte durch die durchschnittliche Distanz des Maßes. Im *soniXplorer* wird jedoch eine andere Strategie verfolgt, die eine genauere Berücksichtigung der verfügbaren Informationen gewährleistet: Anstatt eines einzigen linearen Modells, das alle Abstandsmatrizen kombiniert, wird für jede Teilmenge Metadaten-basierter und kollaborativer Maße ein separater Gewichtsvektor ermittelt, der die Distanzen dieser Auswahl mit den inhaltsbasierten Abständen verknüpft.

Aufgrund der begrenzten Anzahl dieser Teilmengen und der effizienten Gewichtsanzpassung mit Hilfe eines linearen Regressors ist die Erstellung und Wartung dieser Modellvielfalt realisierbar.

Kapitel 3

Explorationsumgebung

Das im letzten Kapitel entwickelte Distanz-Maß für Musikstücke, das durch die Kombination von unterschiedlichen Abstandsberechnungen den facettenreichen Begriff der Ähnlichkeit zu einer singulären Größe zusammenfasst, eröffnet vielfältige Anwendungsmöglichkeiten. So erlaubt es den Aufbau eines Systems zum Query-By-Example-Retrieval oder kann als Basis eines personalisierten Empfehlungssystems fungieren, das aufgrund eines erlernten Benutzer-Profiles ähnliche Stücke identifizieren und präsentieren kann.

Ebenso ist es jedoch als Grundlage einer automatischen Musikkollektions-Strukturierung die Voraussetzung für eine Explorationsumgebung, die dem Benutzer die intuitive Navigation in einer für ihn verständlich organisierten Anordnung ermöglicht. Die ungerichtete aber durch das Ähnlichkeitsmaß angeleitete Suche ohne klar formuliertes Informationsbedürfnis stellt wie in Abschnitt 1.2 dargestellt die häufigere und natürlichere Zugriffsform auf Musiksammlungen dar.

In diesem Kapitel wird auf der Grundlage der beschriebenen Distanzfunktion die Entwicklung einer multimodalen Explorationsumgebung beschrieben, die visuelle Darstellung und akustische Orientierung zu einer immersiven virtuellen Welt kombiniert, in der sich der Benutzer frei bewegen kann. Kapitel 4 diskutiert im Anschluss daran Möglichkeiten der Interaktion des Anwenders mit dieser Umgebung.

Auf eine Beschreibung verwandter Arbeiten zu Musik-Explorationsumgebungen (Abschnitt 3.1) und Methoden zur Topologie-erhaltenden Transformation der Ausgangs-Distanzmatrix in einen Euklidischen Vektorraum (Abschnitt 3.2) folgt in Abschnitt 3.3 eine Diskussion des zentralen Algorithmus zur Positionierung und Abgrenzung von Musikstücken der betrachteten Kollektion auf einer virtuellen Landkarte. Abschnitt 3.4 geht daraufhin in einzelnen Unterkapiteln auf die visuelle und akustische Darstellung der generierten Strukturierung in Form einer navigierbaren dreidimensionalen Landschaft ein, die populären Computer-Spiel-Umgebungen ähnelt.

3.1 Forschungsumfeld

Konzentrierte sich die Music-Information-Retrieval-Forschung zunächst primär auf Analyseverfahren zur Herleitung semantisch höherwertiger musikalischer Konzepte aus dem vorliegenden Audiosignal und die Unterstützung eines gerichteten Retrieval-Prozesses (z.B. in

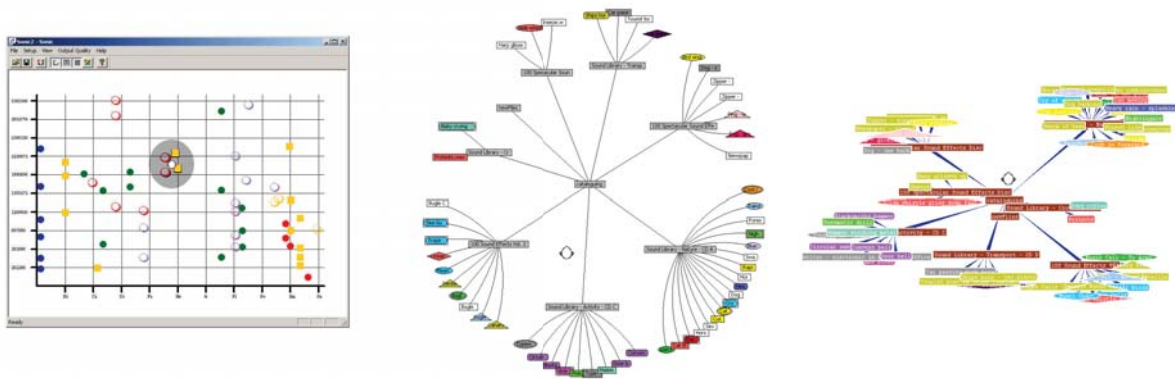


Abbildung 3.2: Sonic-Browser-Visualisierung als Sternfeld-Diagramm, HyperTree und TouchGraph [BF03]

Als letzte Visualisierungstechnik untersuchen Torrens et al. schließlich die Konstruktion einer Tree-Map, die durch sukzessive Aufteilung der Präsentationsfläche in rechteckige Segmente, deren Größe proportional zur Anzahl von Stücken der entsprechenden Kategorie ist, die Zusammensetzung der Kollektion entlang der Hierarchie Genre $\rightarrow$ Untergenre $\rightarrow$ Künstler darstellt (vergl. Abbildung 3.1 (3)).

Alle drei Visualisierungsansätze gruppieren Stücke auf der Basis von assoziierten Metadaten und berücksichtigen somit keine (möglicherweise Genre- und Veröffentlichungsjahr-übergreifende) perzeptuelle Ähnlichkeit zwischen Stücken. Ferner beschränken sich Torrens et al. auf die rein visuelle Präsentation einer persönlichen Musiksammlung. Die akustische Wiedergabe muss in ihren Systemen explizit durch den Benutzer für ein ausgewähltes Stück angestoßen werden. Wie in Abschnitt 1.2.3 dargestellt, verhindert der hieraus resultierende Medienbruch einen schnellen und intuitiven Zugang zur Kollektion.

3.1.2 Sonic Browser [FB01, BFTC02, BF03]

Ein erster Ansatz zur Integration akustischer und visueller Darstellung stammt von Fernström und Brazil. Ihr in [FB01] vorgestellter *Sonic Browser* positioniert Titel anhand einer konfigurierbaren Metadaten-Auswahl in einem zweidimensionalen x-y-Diagramm. Weitere Attribute eines Stückes werden durch Farb- und Formgebung des Symbols visualisiert (vergl. Abbildung 3.2).

Ferner lässt sich ein Avatar in der Darstellung platzieren. Diese Benutzer-Repräsentanz ist umgeben von einer kreisförmigen *Aura* eines konfigurierbaren Radius', die den Bereich des Diagramms abgrenzt, dessen Stücke akustisch wiedergegeben werden. Die relative Position eines Titels zum Standpunkt des Avatars spiegelt sich dabei in der Positionierung des Klangs im akustischen Raum wider.

Brazil und Fernström erweitern ihr System in [BF03] um alternative synchronisierte Visualisierungstechniken, die verschiedene konsistente Sichten auf die gleiche Datenmenge offerieren. Neben der im Sonic Browser realisierten Sternfeld-Darstellung sind dies die im vorherigen Abschnitt genannten TreeMap-Visualisierung sowie eine HyperTree- und eine TouchGraph-Aufbereitung (vergl. Abbildung 3.2). Stets erlaubt die Positionierung eines Benutzer-Avatars

die ergänzende akustische Exploration der generierten Visualisierung, die die Platzierung der Titel als Klangquellen im Raum bestimmt.

Um die Beschränkung auf assoziierte Metadaten zur Organisation der Musikkollektion zu überwinden, integrieren Brazil et al. die Signalverarbeitungs-Toolbox Marsyas in ihr System [BFTC02]. Dies erlaubt z.B. die Anwendung eines automatischen Genre-Klassifikators oder den Zugriff auf inhaltsbasierte Merkmale wie die Stärke des dominierenden Rhythmus-Schlages. Ferner lässt sich die Kollektion mit Hilfe des k -means-Algorithmus in k disjunkte Cluster partitionieren, die gemäß einem inhaltsbasierten Maß zueinander ähnliche Titel zusammenfassen.

Diese Anwendung eines Ähnlichkeitsmaßes bleibt jedoch auf die Zuweisung einer nominalen Cluster-Bezeichnung beschränkt: Zum einen werden keine Beziehungen zwischen den Gruppen repräsentiert, zum anderen kann aus der Darstellung nicht ermittelt werden, wie eindeutig die Zugehörigkeit eines Stückes zu seinem Cluster ist. Damit gibt die von Brazil et al. vorgeschlagene Darstellung die Ähnlichkeitsstruktur der Kollektion nur recht grob wieder.

3.1.3 Islands of Music [PRM02]

Eine der grundlegenden Arbeiten zur Visualisierung von Musikkollektionen stammt von Pampalk et al. und ist unter dem Schlagwort *Islands of Music* bekannt.

Ausgangspunkt ist die Berechnung von inhaltsbasierten Merkmalen, die die klanglichen Charakteristika von Musikstücken modellieren. Pampalk setzt hierzu auf die in Abschnitt 2.1.2 beschriebenen Fluctuation Patterns und fasst diese Fenster-basierte Größe durch den Median der Einzelberechnungen zusammen. Mit Hilfe einer Hauptkomponentenanalyse wird die Dimension des Merkmalsraumes von ursprünglich 1200 auf 80 reduziert.

Ausgehend von dieser Repräsentation der Stücke als 80-dimensionale Feature-Vektoren wird eine selbstorganisierende Merkmalskarte (Self-Organizing Map, SOM) konstruiert [Koh01]. Pampalk et al. verwenden für eine Kollektion aus 359 Titeln eine rechteckige Karte mit 14×10 Einheiten. Resultat dieses unüberwachten Lernprozesses ist die Zuweisung von Titeln zu 2-dimensionalen Gitterpunkten unter einer Topologie-erhaltenden nicht-linearen Abbildung: Stücke, die im hochdimensionalen Raum einen geringen Abstand zueinander haben, werden auch auf der Karte nahe beinander positioniert. Damit ordnet die SOM-Generierung über die Gruppierung ähnlicher Stücken in Gitterpunkten hinaus diese Cluster entsprechend der perzeptuellen Ursprungsdistanz zueinander an und ermöglicht so eine intuitive Visualisierung der Struktur des hochdimensionalen Merkmalsraums. Eine genauere Diskussion des SOM-Algorithmus' findet sich in Abschnitt 3.3.

Mit Hilfe eines robusten Rang-orientierten Schätzverfahrens (*Smoothed Data Histograms*) ermitteln Pampalk et al. aus der Positionierung von Titeln auf SOM-Zellen die Stückdichte der einzelnen Gitterpunkte. Codiert man diese Dichte durch ein geeignetes Farbschema, lässt sich die Verteilung der Titel bei Formulierung einer „Island of Music“-Metapher anschaulich visualisieren: Bereiche der Karte, denen nur wenige Stücke zugeordnet sind, werden als blaues Wasser dargestellt, aus dem sich „Musik-Inseln“ mit unterschiedlichem Vegetationsprofil erheben. Abhängig von der Titeldichte eines Gitterpunktes wird hierzu zwischen blauer (Wasser), gelber (Strand), grüner (Rasen) und braun-weißer (schneebedeckte Berge) Farbgebung unterschieden.

Abbildung 3.3 zeigt ein Beispiel einer so visualisierten selbstorganisierenden Merkmalskarte. Deutlich zu erkennen ist die Zusammenfassung ähnlicher Stücke zu „homogenen Inseln“ (vergl.



Abbildung 3.3: Islands of Music-Visualisierung nach Pampalk [Bau05b]

z.B. die Ansammlung klassischer Stücke in der rechten unteren Ecke) und die Abgrenzung von Gruppen anders klingender Titel durch blaues Wasser.

Auf der Basis dieses Ansatzes sind eine Reihe von Erweiterungen, Varianten und Anwendungen der SOM-Visualisierung von Musikkollektionen entwickelt worden: Hierzu zählt die synchronisierende Kopplung zwischen SOMs unterschiedlicher Ähnlichkeitsmaße, die eine integrierte Betrachtung der Sammlung aus verschiedenen Blickwinkeln ermöglicht [PDW03a]. Ferner beschäftigen sich Rauber et al. in [DMR00] und [RPM02] mit Fragen der Skalierbarkeit des Ansatzes und entwickeln eine hierarchische Erweiterung von Merkmalskarten. Neumayer et al. untersuchen in [NDR05], inwiefern sich SOMs zur intuitiven Zusammenstellung von Wiedergabelisten eignen.

All diese Ansätze beschränken sich im Wesentlichen auf die visuelle Beschreibung einer Musiksammlung und ignorieren damit die Chancen der Unmittelbarkeit einer akustischen Explorationserfahrung.

3.1.4 MusicMiner [MUNS05]

Eine mit Pampalks *Islands of Music* stark verwandte Explorationsumgebung, die ebenfalls auf einer Strukturierung mit Hilfe selbstorganisierender Merkmalskarten basiert, wurde von Mörchen unter der Bezeichnung *MusicMiner* entwickelt [MUNS05].

Grundlage der Landkarten-Generierung ist die Charakterisierung der zu visualisierenden Stücke durch die in Abschnitt 2.8.1 bereits genannte Merkmalsmenge, die sich aus 20 aufwändig selektierten inhaltsbasierten Features zusammensetzt. Wie in [MUTL06] berichtet haben sich diese Merkmale als besonders diskriminativ hinsichtlich des Genres des Titel erwiesen.

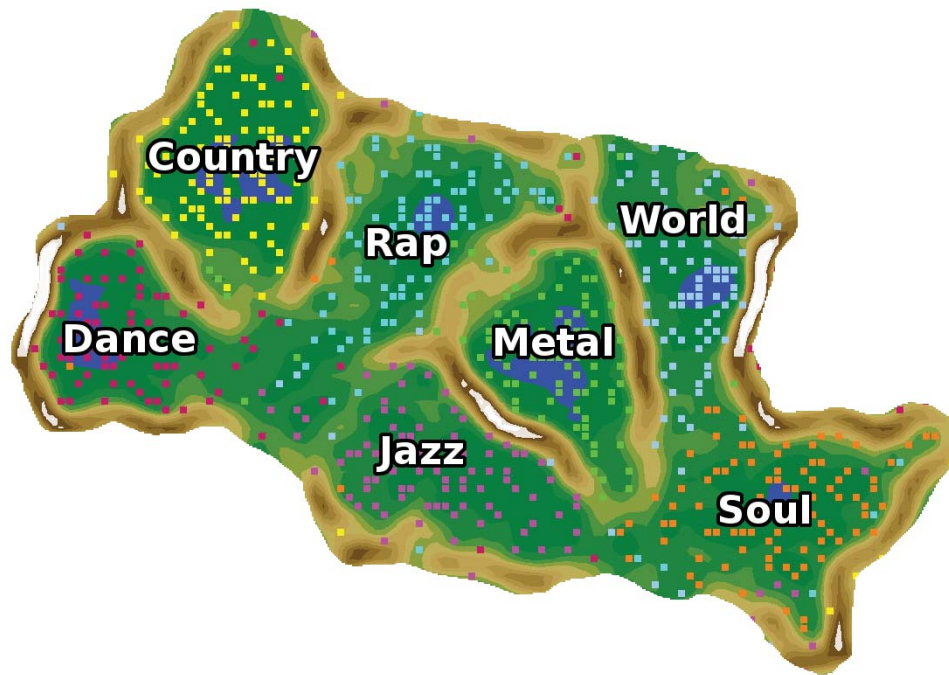


Abbildung 3.4: Visualisierung einer Musikkollektion mit Hilfe des MusicMiners [MUNS05]

Im Gegensatz zur vergleichsweise kleinen Anzahl Zellen, die in Pampalks SOMs Verwendung finden, setzen Mörchén et al. auf sog. *emergente* Merkmalskarten [Ult92], die sich durch eine deutliche höhere Neuronen-Dichte auszeichnen. Dies ermöglicht nicht nur ein Clustering von Songs zu Gruppen, deren Abstände zueinander den Verhältnissen im hochdimensionalen Raum entsprechen (inter-cluster distances), sondern erlaubt darüber hinaus die Darstellung von Mikro-Strukturen innerhalb dieser Gruppen (intra-cluster distances). Aus dieser fein aufgelösten Detail-Struktur entsteht („emergiert“) schließlich die Makro-Struktur der betrachteten Kollektion.

Zur Vermeidung der vermehrten Positionierung von Stücken an den Ecken einer rechteckigen Karte [Ult03a] verknüpfen Mörchén et al. die Enden zu einem geschlossenen Ring (Toroid), in dem alle Zellen die identische Anzahl von Nachbarn besitzen.

Augenscheinlichster Unterschied zur Visualisierung von Pampalk ist die Methode zur Darstellung der SOM-Struktur. An die Stelle einer Schätzung der Stückdichte für einen Gitterpunkt tritt die Berechnung des durchschnittlichen Abstandes des Modellvektors der Zelle zu den umgebenden Modellvektoren. Hohe Werte stehen damit für eine klare Abgrenzung des Kartenbereichs von seiner Nachbarschaft.

Interpretiert man diese so berechnete *U-Matrix* [Ult92] als Höhenprofil, so ergibt sich eine den Islands of Music entgegengesetzte Visualisierung: Zusammenhänge Gruppen ähnlicher Stücke befinden sich in flachen Tälern, die durch Hügel und Bergketten separiert werden.

Abbildung 3.4 zeigt ein Beispiel einer solchen MusicMiner-Karte, wie sie aus einer Entfaltung der Toroid-SOM und der Interpretation der U-Matrix durch ein geeignetes Farbschema entsteht. Die nur zur Evaluation verwendeten Genre-Bezeichner der platzierten Songs lassen

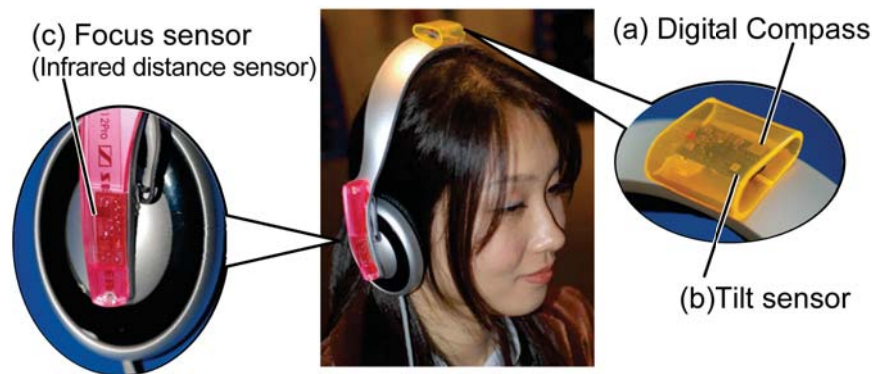


Abbildung 3.5: Angebrachte Kopfhörer-Sensoren zur Detektion von Kopf-Bewegung und Fokussierungs-Geste nach [HL06]

allein aufgrund inhaltsbasierter Merkmale generierte deutlich abgegrenzte homogene Bereiche klar erkennen und illustrieren ferner die Qualität der MusicMiner-Feature-Menge zur Genre-Diskrimination.

Wie den Islands of Music fehlt dem MusicMiner-System jedoch eine akustische Orientierungshilfe für den Benutzer, die über die übliche Multimedia-Player-Funktionalität der expliziten Wiedergabe eines ausgewählten Stückes hinausgeht.

3.1.5 Akustische Exploration [HL06]

Im Gegensatz dazu stellen Hamanaka und Lee ein System vor, das den Anwender mit Hilfe eines speziellen Kopfhörers in eine virtuelle akustische Umgebung versetzt, in der er die gegebene Musiksammlung allein auf der Grundlage räumlich positionierter Klangquellen ohne unterstützende Visualisierung explorieren kann.

Eine intuitive Navigation in dieser virtuellen Umgebung erfordert eine enge Kopplung zwischen der Bewegung des Benutzers und der Anpassung der akustischen Wiedergabe. Hierzu wird ein handelsüblicher Kopfhörer durch drei Sensoren ergänzt (vergl. Abbildung 3.5):

1. Ein digitaler Kompass erfasst die Blickrichtung der Anwenders in der Horizontalebene.
2. Ein weiterer Detektor misst die Neigung des Kopfes und damit den Blickwinkel in der Medianebene.
3. Ein Infrarot-Abstands-Sensor erkennt das Heranführen einer Hand zur Ohrmuschel.

Hamanaka und Lee gehen von einer stationären Positionierung des Anwenders im virtuellen Raum aus. Um diesen Standpunkt herum platzieren sie gemäß einem konfigurierbaren Anordnungsmuster bis zu zehn permanent erklingende Musikstücke, die auf der Grundlage obiger Sensoren in ihrer Lautstärke angepasst und spatialisiert werden.

Abhängig von der Entfernung zum Benutzer und dem Winkel zwischen Blickrichtung und Vektor zur Titel-Position wird ein Stück mit unterschiedlichem Pegel auf linkem und rechtem

Kanal wiedergegeben. Hierdurch entsteht beim Hörer die Illusion einer stabilen Lokalisation der Klangquelle am festgelegten Ort im virtuellen Raum trotz einer Bewegung des Kopfes.

Neigt der Benutzer den Kopf etwas nach unten, treten Stücke, die sich in geringerer Distanz zum Standpunkt befinden, stärker in den Vordergrund. Analog erlaubt der Neigungs-Sensor die Hervorhebung von weiter entfernten Titeln, die ansonsten durch davor platzierte Stücke verdeckt wären.

Eine intuitive Bewegung zur Abschattung irrelevanter Klangquellen in natürlichen Umgebungen ist das Heranführen der geschlossenen Hand zur Ohrmuschel, um die Direktionalität des Hörbereichs zu verstärken. Diese Geste detektieren Hamanaka und Lee durch den letzten oben angeführten Sensor und heben in Abhängigkeit von der gemessenen Distanz zwischen Kopfhörerohr und Hand solche Titel hervor, die sich im Zentrum der Blickrichtung des Benutzers befinden. Diese Fokussierung erlaubt es dem Anwender, einzelne Quellen aus dem Klanggemisch zu trennen und genauer zu betrachten.

Die realisierte Spatialisierung von Klangquellen beschränkt sich bei Hamanaka und Lee auf eine richtungsspezifische Verteilung des Signals auf den linken und rechten Kopfhörer-Kanal. Die Lokalisation der Schallquelle ergibt sich für den Benutzer damit allein aufgrund der interauralen Pegeldifferenz. Wie in Abschnitt 3.4.2 näher ausgeführt sind bei der Ortung von Klangquellen in natürlichen Umgebungen jedoch weitere Einflussfaktoren wie interaurale Laufzeitdifferenzen oder klangliche Abschattungseffekte durch Torso und Kopf von Bedeutung. Erst die Berücksichtigung dieser Größen ermöglicht eine Unterscheidung zwischen Klangquellen vor und hinter der Hörposition.

Die von Hamanaka und Lee genutzten Sensoren detektieren lediglich die Blickrichtung und verlangen daher die stationäre Platzierung des Anwenders im virtuellen Raum. Aus dieser Einschränkung resultiert die geringe Anzahl von maximal zehn Stücken, die sich ohne Überforderung des Hörers gleichzeitig präsentieren lassen.

3.1.6 Dreidimensionale multimodale Explorationsumgebung [KPSW07b]

Knees beschreibt in [KPSW07b] unter der Bezeichnung *nepTune* einen unabhängig vom *soniXplorer* entwickelten Prototypen, der SOM-basierte Visualisierung und akustische Orientierung kombiniert.

Ausgangspunkt ist die Topologie-erhaltende Anordnung von Stücken einer Musikkollektion auf einer selbstorganisierenden Merkmalskarte gemäß dem Island-Of-Music-Ansatz von Pampalk (vergl. Abschnitt 3.1.3). Wie dort beschrieben erlaubt die Platzierung von Titeln auf Gitterpunkten der Karte eine Schätzung der Song-Dichte pro Zelle. Zur 2-dimensionalen Visualisierung ordnet Pampalk jeder Dichte eine Farbgebung zu, die ein Höhenprofil der planaren Karte suggeriert.

Knees et al. verwenden diese Höheninformation, um eine dreidimensionale virtuelle Landschaft aus Bergen unterschiedlicher Größe zu generieren, die sich aus dem umgebenden Ozean abhebt. Abbildung 3.6 zeigt ein Beispiel einer so visualisierten SOM.

Auch wenn die dreidimensionale Illustration des Höhenprofils keine über die Islands-Of-Music-Darstellung hinausgehenden Informationen beinhaltet, so erlaubt diese Darstellung doch eine intuitivere Erfassung der Kartenstruktur und trägt in Verbindung mit einer einfachen Game-Controller-basierten Navigation zum Eindruck einer immersiven virtuellen Welt bei.

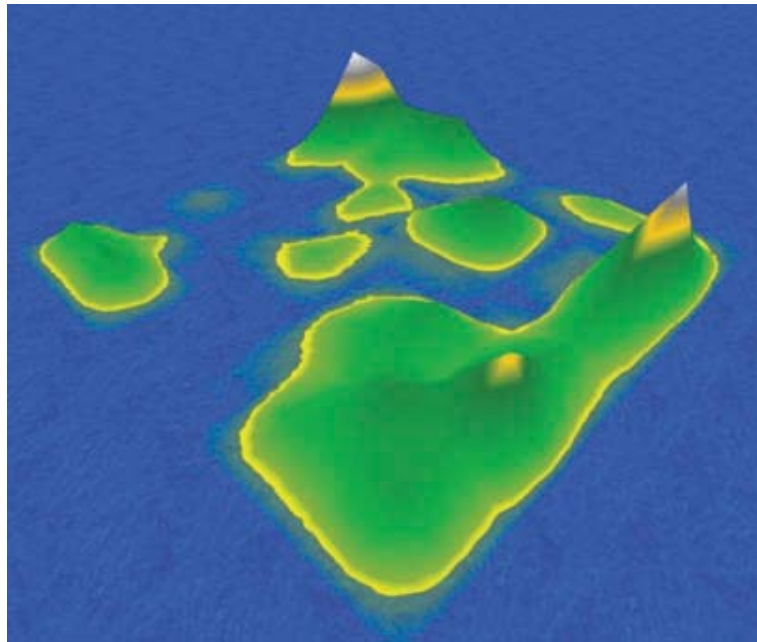


Abbildung 3.6: Visualisierung einer Musikkollektion als dreidimensionale Landschaft [KPSW07b]

Diese Illusion wird unterstützt durch die simultane akustische Wiedergabe der Stücke in unmittelbarer Umgebung des aktuellen Standpunktes. In Abhängigkeit von der relativen Position der Klangquelle weisen Knees et al. das Audiosignal entsprechenden Kanälen eines 5.1-Soundsystems zu und ermöglichen hierdurch die akustische Ortung von Songs in geringer Distanz.

Neben einer Beschriftung von Kartenregionen mit ID3-basierten Metadaten der platzierten Stücke erlaubt *nepTune* die automatische Annotation mit beschreibenden Termen eines vorgegebenen musikalischen Vokabulars. Hierzu werden die Resultate Künstler-spezifischer Google-Anfragen ähnlich den Verfahren zur Berechnung textbasierter Ähnlichkeitsmaße (vergl. Abschnitt 2.4.3) analysiert. Alternativ nutzen Knees et al. die Ergebnisse einer kontextuellen Google-Bildersuche zur optischen Kennzeichnung von Kartenregionen. Abbildung 3.7 stellt diese drei Anzeige-Modi einer unbeschrifteten Landschaftsvisualisierung gegenüber.

Auch wenn *nepTune* als multimodale immersive Explorationsumgebung betrachtet werden kann und viele Gemeinsamkeiten mit dem im Rahmen dieser Arbeit entwickelten *soniXplorer*-Prototypen teilt, sind dennoch eine Reihe von Schwächen des Ansatzes von Knees et al. zu benennen:

Die akustische Wiedergabe verlangt die Präsenz eines 5.1-Soundsystems. Dies verhindert eine Portierung der Explorationsumgebung auf populäre portable Musikplayer. *soniXplorer* dagegen setzt auf HRTF-basierte Spatialisierung von Klangquellen.

Knees et al. berichten, dass verschiedene Nutzer sich im Rahmen einer qualitativen Evaluation des Systems eine Vergrößerung der Landkarte wünschten (obwohl im Szenario nur maximal 50 Stücke platziert wurden), um die akustische Orientierung in dichten Regionen zu vereinfachen. Diese Reaktion ist Zeichen für die Notwendigkeit einer Funktionalität zur

Fokussierung der akustischen Wiedergabe wie sie etwa von Hamanaka und Lee im vorigen Abschnitt vorgeschlagen wurde und im *soniXplorer* realisiert ist.

3.1.7 Zusammenfassung und Gegenüberstellung

Tabelle 3.1 fasst die Beschreibungen der in den vorangehenden Abschnitten vorgestellten Explorationsumgebungen hinsichtlich verschiedener Kriterien zusammen und stellt die Merkmale der Systeme übersichtlich gegenüber.

Modalität der Präsentation

Das erste differenzierende Kriterium betrifft die Modalität der Präsentation. Die Mehrzahl der publizierten Arbeiten beschränkt sich auf eine visuelle Darstellung der Musikkollektion. Eine Ausnahme bildet das Kopfhörer-basierte System von Hamanaka und Lee zur rein akustischen Exploration. Erste Ansätze zur Integration von visueller und akustischer Darstellung sind im *SonicBrowser* und im dreidimensionalen *nepTune*-Prototypen realisiert.

Eine Arbeitsthese der *soniXplorer*-Entwicklung ist die Notwendigkeit einer solchen *multimodalen* Präsentation der Musikkollektion, um einen Medienbruch aufgrund einer rein visuellen Darstellung musikalischen (und damit primär akustischen) Materials zu vermeiden und dem Benutzer die Möglichkeit zur informierteren Navigation im System zu ermöglichen.

Strukturierungs-Grundlage

Ferner unterscheiden sich die vorgestellten Explorationsumgebungen in der Datenbasis, die zur Strukturierung der Stücksammlung verwendet wird. Während die in Abschnitt 3.1.1 beschriebenen Visualisierungsverfahren allein assoziierte Metadaten betrachten, führen der *Islands-of-Music*-Ansatz, *MusicMiner* sowie der *nepTune*-Prototyp eine inhaltsbasierte Ähnlichkeitsanalyse der Musikkollektion durch, in der mit Hilfe von Signalverarbeitungsalgorithmen perzeptuelle Ähnlichkeiten zwischen Stücken ermittelt werden. Hierzu stützen sich die Systeme von Pampalk (*Islands of Music*) und Knees (*nepTune*) auf die in Abschnitt 2.1.2 diskutierten *Fluctuation Patterns*, Mörchen et al. (*MusicMiner*) verwenden eine in einem aufwändigen Selektionsprozess ausgewählte Genre-diskriminative Feature-Kombination (vergl. Abschnitt 2.8.1).

Die Verwendung von paarweisen automatisch berechneten Distanzen gegenüber einer ausschließlich auf Metadaten basierenden Kategorisierung weist die folgenden Vorteile auf:

- Metadaten bedürfen einer aufwändigen manuellen Erstellung und Pflege. Die Anwendung von Signalverarbeitungsalgorithmen entfernt die Abhängigkeit von menschlicher Interaktion und skaliert besser.
- Das Ähnlichkeitsempfinden des Benutzers, das für eine intuitive Navigation in einer Explorationsumgebung erforderlich ist, wird durch die Ähnlichkeit in den Metadaten der Stücke nur unzureichend abgebildet. So ist z.B. das von Torrens et al. zur Strukturierung verwendete Veröffentlichungsjahr eines Titels nur bedingt zur Beschreibung perzeptueller Ähnlichkeit nützlich; Metadaten wie z.B. der Autor von Stücken erlaubt keinerlei sinnvolle Abstands-Berechnung.

- Metadaten-basierte Klassifikationen (z.B. die Eingruppierung in Genre-Klassen) stellen eine übergroße Vereinfachung der Ähnlichkeitsräume dar. Zum einen gestatten sie keine feingranularere Bestimmung der Struktur von Titeln innerhalb einer Klasse (intra-class distance), zum anderen lassen sie sich untereinander nicht vergleichen (inter-class distance).

Aus diesen Gründen werden im *soniXplorer* paarweise Abstände zwischen Stücken gemäß einem aufwändig kombinierten Distanz-Mix berechnet und als Grundlage der Strukturierung der Kollektion zur Exploration verwendet.

Eine Sonderrolle nimmt erneut die rein akustische Explorationsumgebung von Hamanaka und Lee ein: Die Anordnung von Stücken erfolgt hier nach einem Benutzer-konfigurierbaren (aber Kollektions-unabhängigen) räumlichen Schema.

Strukturierungs-Technik

Auf der Basis der gewählten Basisdaten werden unterschiedliche Verfahren zum Aufbau einer explorierbaren Struktur angewendet. Während die Metadaten-basierte Visualisierung ähnlich dem *SonicBrowser* Standard-Visualisierungsmethoden des Data Mining nutzen, bilden Pampalk und Knees die hochdimensionale Ähnlichkeitsstruktur der Stückkollektion Topographie-erhaltend mit Hilfe selbstorganisierender Karten (vergl. Abschnitt 3.3) auf eine zweidimensionale Landschaft ab. Im Island-of-Music-Ansatz wie auf bei *nepTune*-Prototypen ähnelt diese Dimensionsreduktion aufgrund der relativ geringen Neuronendichte einem k-Means-Clustering mit dem Unterschied, dass die SOM-Generierung die detektierten Cluster zueinander in Beziehung stellt. Mörchen et al. verwenden anstatt dessen emergente selbstorganisierende Karten, die sich durch eine deutlich höhere Zellen-Anzahl auszeichnen und so auch die Mikro-Struktur innerhalb von Stückgruppen offenbaren.

Aufgrund der in Abschnitt 3.1.4 diskutierten Vorteile dieser emergenten SOMs findet diese Technik auch im *soniXplorer* Anwendung.

SOM-Visualisierungs-Metapher

Eine generierte selbstorganisierende Karte lässt sich auf verschiedene Weisen visualisieren. Während Pampalk und Knees die einem Neuron zugeordnete Dichte schätzen und Zellen mit hoher Anzahl zugeordneter Titel eine große Landschaftshöhe zuweisen, verfolgen Mörchen et al. eine dazu entgegengesetzte Visualisierungsmetapher, in der auf der Grundlage der Verschiedenheit eines Neurons gegenüber seiner direkten Umgebung homogene Kartenbereiche als tiefe Täler dargestellt werden, die durch hohe Bergketten separiert werden. Im Island-of-Music-Ansatz wird ebenso wie im *MusicMiner* diese Höheninformation durch eine suggestive Farbgebung visualisiert. Knees dagegen entwickelt eine immersive dreidimensionale virtuelle Welt, in der der Benutzer mit Hilfe eines Game-Controllers navigieren kann.

Im *soniXplorer* ist eine Kombination der Ansätze von Mörchen und Knees implementiert: Die Visualisierungsmetapher stammt aus dem *MusicMiner* (homogene Stück-Regionen als Täler, Clustergrenzen als Bergketten). Wie Knees wird das berechnete Höhenprofil jedoch zum Aufbau einer dreidimensionalen Landschaft verwendet.

Akustische Wiedergabe

Die letzte in Tabelle 3.1 aufgeführte Kategorie zur Unterscheidung der diskutierten verwandten Ansätze betrifft die akustische Präsentation der Stücke während des Explorationsprozesses und ist daher nur für die Systeme relevant, die eine direkte Wiedergabe der Titel in die Exploration integrieren: den SonicBrowser, die rein akustische Exploration nach Hamanaka und Lee sowie den nepTune-Prototypen.

Die ersten beiden dieser Systeme realisieren eine einfache Verteilung des Tonsignals auf die beiden verfügbaren Stereo-Kanäle des Hörers in Abhängigkeit von der relativen Position der Klangquelle zum Betrachter. Im SonicBrowser wird die Wiedergabe auf die Titel in der Umgebung (der *Aura*) des Anwenders beschränkt, im System von Hamanaka und Lee ist die Anzahl der gleichzeitig spielenden Stücke bereits durch die bewusste Platzierung einer begrenzten Anzahl von Titeln im Raum beschränkt. Zur weiteren Eingrenzung dieser Menge implementiert dieses System einen Fokussierungsmechanismus, der eine akustische Konzentration auf frontal positionierte Titel ermöglicht.

An die Stelle einer Stereo-Spatialisierung tritt im nepTune-Prototypen die Wiedergabe auf einem 5.1-Surround-System. Wie in Abschnitt 3.1.6 beschrieben erschwert das Fehlen eines Fokussierungsmechanismus jedoch die akustische Orientierung und führt zu einer Reizüberflutung beim Hörer, die eine Differenzierung der simultan klingenden Stücke verhindert.

Mit Hinblick auf eine spätere Nutzung der Explorationsumgebung im Kontext mobiler Systeme wird im soniXplorer eine virtuelle räumliche Positionierung von Klangquellen mit Hilfe von HRTF-Funktionen für eine binaurale Wiedergabe realisiert (vergl. Abschnitt 3.4.2). Ferner ist ein aufwändiger Fokussierungsmechanismus implementiert.

Fehlende Personalisierung

Keines der in den vorangehenden Abschnitten vorgestellten Explorationsumgebungen erlaubt eine Personalisierung der Landschaft durch den Benutzer, etwa durch eine Repositionierung von Stücken auf der Karte, eine Veränderung des Höhenprofils oder die Platzierung von Orientierungspunkten zur Kennzeichnung von Regionen. In allen Ansätzen ist die Interaktion der Benutzers mit dem System auf ein Navigieren in der Landschaft beschränkt.

3.2 Vektorraum-Repräsentation

Aufgrund der bewussten Integration von Abstandsmaßen, die auf einer paarweisen Berechnung der Unähnlichkeit zweier Stücke basieren, resultiert die in Abschnitt 2.8 beschriebene Kombination verschiedener Einzelmaße zur Berücksichtigung unterschiedlicher musikalischer Teilaspekte in einer quadratischen Distanzmatrix, deren Zeilen- und Spaltenanzahl der Stückmenge entspricht. Die Symmetrie der Einzel-Distanzen überträgt sich auf die linear kombinierte Gesamt-Distanz. Daher reicht die Speicherung einer oberen Dreiecksmatrix aus.

Die Erstellung einer selbstorganisierenden Merkmalskarte setzt jedoch die Repräsentation der zu platzierenden Objekte als Vektoren eines Euklidischen Raumes voraus. Vor einer Anwendung dieses Verfahren ist daher ein Transformationsschritt erforderlich, der jedem Stück einen Merkmalsvektor zuordnet, so dass die Euklidische Distanz zwischen diesen Vektoren

Kriterium	Metadaten-basierte Visualisierung Abschnitt 3.1.1	Sonic Browser Abschnitt 3.1.2	Islands of Music Abschnitt 3.1.3	MusicMiner Abschnitt 3.1.4	Akustische Exploration Abschnitt 3.1.5	nepTune Abschnitt 3.1.6
Modalität der Präsentation	visuell	visuell + akustisch	visuell	visuell	akustisch	visuell + akustisch
Strukturierungs-Grundlage	Metadaten	Metadaten (+ inhaltsbasierte Klassifikation)	Fluctuation Patterns (inhaltsbasierte Ähnlichkeit)	MusicMiner-Merkmale (inhaltsbasierte Ähnlichkeit)	Vorgegebene Positionierung	Fluctuation Patterns (inhaltsbasierte Ähnlichkeit)
Strukturierungs-Technik	Disc-View, Rectangle-View, Tree-Map	Stemenfeld-Diagramm, HyperTree, TouchGraph	SOM	Emergente SOM	Vorgegebene Positionierung	SOM
SOM-Visualisierungs-Metapher	/	/	Cluster als Inseln im Meer, Trennung durch Wasser	Stücke in Tälern, Clustergrenzen als Bergketten	/	Cluster als Inseln im Meer, Trennung durch Wasser 3-D Darstellung
Akustische Wiedergabe	/	Stereo-Spatialisierung aller Stücke in Aura	/	/	Stereo-Spatialisierung, Fokussierungs-Mechanismus	5.1-Wiedergabe der Umgebung

Tabelle 3.1: Gegenüberstellung der diskutierten verwandten Arbeiten zur Exploration von Musiksammlungen

weitestmöglich dem Abstand in der Ursprungs-Distanzmatrix entspricht (vergl. Abbildung 1.11).

Diese Problemstellung ist ein typisches Anwendungsszenario des *Multidimensional Scaling (MDS)* [dL01], das als Optimierungsverfahren verstanden werden kann, dessen Ziel die Minimierung der Abweichung paarweiser *Unähnlichkeiten* (dissimilarities) von der Euklidischen *Distanz* (distance) zugeordneter Vektor-Repräsentationen der Objekte darstellt.

Für eine Menge von N Titeln und eine vorgegebene Ziel-Dimensionalität k bezeichne im folgenden $\vec{x}_i$ den k -dimensional Vektor, der Stück i zugeordnet wird. $\delta_{i,j}$ beschreibe die Unähnlichkeit zwischen Stück i und j , die zu approximieren ist.

Mit dieser Notation lässt sich ein erstes einfaches Optimierungskriterium formalisieren:

$$\text{Stress}_1(\vec{x}_1, \dots, \vec{x}_N) = \sum_{\substack{i,j=1\dots N \\ i \neq j}} (\delta_{i,j} - \|\vec{x}_i - \vec{x}_j\|)^2$$

Die Fülle verschiedener MDS-Ansätze lässt sich wie folgt klassifizieren:

1. *Klassisches MDS (auch classical Torgerson-Gower inner product scaling)* [Tor52]

Nach einer geeigneten Normierung der $\delta_{i,j}$ und unter der Vorgabe, dass die komponentenweisen Mittelwerte der $\vec{x}_i$ 0 sind, lässt sich ein als *Strain* bezeichnetes Optimierungskriterium formulieren, das auf einem Skalarprodukt-Vergleich basiert:

$$\text{Strain}(\vec{x}_1, \dots, \vec{x}_N) = \left(\frac{\sum_{\substack{i,j=1\dots N \\ i \neq j}} (\tilde{\delta}_{i,j} - \langle \vec{x}_i, \vec{x}_j \rangle)^2}{\sum_{\substack{i,j=1\dots N \\ i \neq j}} \langle \vec{x}_i, \vec{x}_j \rangle^2} \right)^{1/2}$$

Hierbei stellt $\tilde{\delta}_{i,j}$ die normierte Unähnlichkeit zwischen Musiktitel i und j dar. Diese Formalisierung erlaubt die Rückführung der Optimierung auf ein effizient und geschlossen lösbares Eigenwertproblem [dL01].

Wie in [BSL⁺96] beschrieben, ist eine Erweiterung des klassischen MDS-Ansatzes um eine Gewichtung von Abweichungen, die die Relevanz der Approximation einzelner Unähnlichkeiten widerspiegeln könnte, kaum möglich. Auch die schwächere Forderung nach einer Optimierungsprozedur, die mit fehlenden $\delta_{i,j}$ umzugehen weiß, ist nicht erfüllt. Wie in Abschnitt 6.2 erläutert wird, ist dies eine notwendige Voraussetzung für eine Skalierbarkeit der Distanzberechnung im **soniXplorer**.

Um eine graduelle Anpassung der Explorationsumgebung nach der Aktualisierung der Unähnlichkeitsmatrix aufgrund von Benutzer-Feedback zu ermöglichen, ist die Vorgabe eines Anfangszustandes, der als Startpunkt einer Optimierung dient, sinnvoll. Auch dies leistet das klassische MDS nicht.

2. *Nicht-Metrisches MDS* [Kru64a, Kru64b, She62]

Beschränkt man sich auf die aus den $\delta_{i,j}$ ableitbare Ähnlichkeits-Rangfolge und sucht eine Zuordnung von $\vec{x}_i$, die die Stücke ordnungserhaltend im k -dimensionalen Raum positioniert, lassen sich Verfahren des Nicht-Metrischen MDS verwenden.

Vor einer Suche nach geeigneten $\vec{x}_i$ wird ein durch isotonische Regression bestimmter monotoner Zusammenhang zwischen den Unähnlichkeiten $\delta_{i,j}$ und ihrer *Disparität* (disparity) ermittelt. Auf der Grundlage dieses Distanzsurrogates werden schließlich nach ähnlichen Kriterien wie beim metrischen MDS Stress-minimierende Vektor-Repräsentationen x_i für alle Stücke i gesucht.

Da die verwendeten perzeptuellen Unähnlichkeits-Maße im *soniXplorer* jedoch tatsächlich metrisch und nicht nur ordinal skaliert sind, reduziert die Anwendung eines Rang-basierten Verfahrens die verfügbare Ähnlichkeits-Information unangemessen.

3. *Metrisches MDS (auch Kruskal-Shepard distance scaling)* [Kru64a, Kru64b]

Metrisches MDS kann als Verallgemeinerung des klassischen MDS betrachtet werden. Die Generalisierung erlaubt eine einfache Gewichtung einzelner Abweichungen $(\delta_{i,j} - \|\vec{x}_i - \vec{x}_j\|)^2$ im Gesamt-Optimierungsziel und ermöglicht somit auch, nur bestimmte Unähnlichkeiten zu berücksichtigen. Diese im *soniXplorer* benötigte Erweiterung führt jedoch dazu, dass iterative Lösungsverfahren an die Stelle eines geschlossen lösbaren Eigenwertproblems treten müssen.

Neben dem oben angeführten Stress₁-Kriterium sind weitere Verlustfunktionen Beschreibung des Optimierungsziels populär. Eine erste Variante normalisiert die summierten Abweichungen durch das Quadrat der Ursprungsunähnlichkeiten

$$\text{Stress}(\vec{x}_1, \dots, \vec{x}_N) = \sqrt{\frac{\sum_{\substack{i,j=1\dots N \\ i \neq j}} (\delta_{i,j} - \|\vec{x}_i - \vec{x}_j\|)^2}{\sum_{\substack{i,j=1\dots N \\ i \neq j}} \delta_{i,j}^2}}$$

Die Wurzelbildung betont kleine Werte dieser Verlustfunktion überproportional, besitzt aufgrund der Monotonie von $\sqrt{\cdot}$ jedoch keinen Einfluss auf das Optimierungsergebnis.

Unter der Bezeichnung *Sammon's Mapping* ist ein MDS-Verfahren bekannt, das auf folgendem Kriterium basiert:

$$\text{Stress}_{\text{Sammon}}(\vec{x}_1, \dots, \vec{x}_N) = \sqrt{\frac{1}{\sum_{i < j} \delta_{i,j}} \sum_{i < j} \frac{(\delta_{i,j} - \|\vec{x}_i - \vec{x}_j\|)^2}{\delta_{i,j}}}$$

Um der Beibehaltung großer Unähnlichkeiten eine höhere Bedeutung beizumessen, lassen sich $\delta_{i,j}$ und $\|\vec{x}_i - \vec{x}_j\|$ in Stress durch ihr Quadrat substituieren:

$$\text{Stress}_{\text{squared}}(\vec{x}_1, \dots, \vec{x}_N) = \sqrt{\frac{\sum_{\substack{i,j=1\dots N \\ i \neq j}} (\delta_{i,j}^2 - \|\vec{x}_i - \vec{x}_j\|^2)^2}{\sum_{\substack{i,j=1\dots N \\ i \neq j}} \delta_{i,j}^4}}$$

Wählt man obige Stress-Funktion als Verlustmaß, so lässt sich daraus ein Kriterium ableiten, das implizit über die Skalierung der $\vec{x}_i$ optimiert [BSL⁺96]:

$$\min_t \text{Stress}(t \cdot \vec{x}_1, \dots, t \cdot \vec{x}_N) = \sqrt{1 - \frac{(\sum_{\substack{i,j=1\dots N \\ i \neq j}} \delta_{i,j} \cdot \|\vec{x}_i - \vec{x}_j\|)^2}{\sum_{\substack{i,j=1\dots N \\ i \neq j}} \delta_{i,j}^2 \cdot \sum_{\substack{i,j=1\dots N \\ i \neq j}} \|\vec{x}_i - \vec{x}_j\|^2}}$$

Eine genauere Betrachtung zeigt, dass sich der Bruch als Cosinusmaß interpretieren lässt, das den Vektor der Unähnlichkeiten $\delta_{i,j}$ mit dem Vektor der entsprechenden Distanzen $\|\vec{x}_i - \vec{x}_j\|$ vergleicht. Die Verlustfunktion quantifiziert damit den Sinus des Winkels zwischen diesen beiden Vektoren und beschreibt somit anschaulich deren Verschiedenheit.

Da obige Verlustfunktionen die Abbildungsfehler einzelner Distanzen lediglich aufsummieren, kann man eine direkte Erweiterung um eine Gewichtung dieser Einzelfehler formulieren. So lässt sich der Stress-Begriff für gefundene Zielvektoren $\vec{x}_1, \dots, \vec{x}_N$ bei gegebenen Gewichten $w_{i,j}$ wie folgt verallgemeinern:

$$\text{Stress}_{\text{weighted}}(\vec{x}_1, \dots, \vec{x}_N) = \sqrt{\frac{\sum_{\substack{i,j=1\dots N \\ i \neq j}} w_{i,j} (\delta_{i,j} - \|\vec{x}_i - \vec{x}_j\|)^2}{\sum_{\substack{i,j=1\dots N \\ i \neq j}} w_{i,j} \delta_{i,j}^2}}$$

Setzt man $w_{i,j} = 0$, so wird die Unähnlichkeit $\delta_{i,j}$ zwischen den Stücken i und j bei der Suche nach geeigneten $\vec{x}_i$ und $\vec{x}_j$ nicht berücksichtigt.

Die Differenzierbarkeit obiger Verlustfunktionen ermöglicht die Anwendung klassischer iterativer Optimierungsverfahren wie Gradient Descent zur Bestimmung lokaler Minima. Für das $\text{Stress}_{\text{weighted}}$ -Maß existiert darüber hinaus ein spezielles Verfahren (SMACOF, [dL77]), das eine schnellere Konvergenz verspricht. Dieser Algorithmus wird auch von der Matlab-Implementierung angewendet, die **soniXplorer** nutzt. Um die Gefahr der Fixierung auf lokale Minima zu verringern, wird das beste Ergebnis aus einer Reihe von MDS-Wiederholungen mit unterschiedlichen Anfangskonfigurationen gewählt.

Alternative Verfahren, die insbesondere eine intuitive Visualisierung der iterativen Anpassung der $\vec{x}_i$ ermöglichen, basieren auf einer physikalischen Simulation, die die Objekte als Punkte im Raum versteht, die durch Federn mit jedem anderen Punkt verbunden sind. Die Kraft einer Feder zwischen i und j ergibt sich dabei aus der Abweichung der Distanz im Raum von der Unähnlichkeit in der Ursprungsmatrix $(\delta_{i,j} - \|\vec{x}_i - \vec{x}_j\|)^2$. Summiert man die gerichteten Einzelkräfte der Federn auf, lässt sich das Objekt im Raum (unter Berücksichtigung einer Dämpfungskonstante) beschleunigen. Gasser et al. nutzen dieses Feder-Modell zur zweidimensionalen Anordnung von Stücken und Audio-Strömen [GFW08].

Die beschriebenen iterativen Verfahren haben eine Laufzeitkomplexität von $\mathcal{O}(N^2)$. Im Rahmen der in dieser Arbeit durchgeführten Experimente erwies sich dies bei einer Kollektionsgröße von bis zu 500 Stücken nicht als problematisch. Für umfangreichere Sammlungen ist möglicherweise auf Algorithmen zurückzugreifen, die auf Kosten der Abbildungsqualität die Laufzeitkomplexität reduzieren (Morrison [MRC03]: $\mathcal{O}(N\sqrt{N})$, Fastmap [FL95, CKGB02]: $\mathcal{O}(N)$).

Multidimensional Scaling kann für $k = 2$ oder $k = 3$ direkt zur Positionierung von Stücken in einer Explorationsumgebung angewendet werden. So beschreiben Lamere und Eck eine interaktive Visualisierung, die Musiktitel in einem dreidimensionalen Raum als Kugeln unterschiedlicher Farbe darstellt (vergl. Abbildung 3.8 [LE07]).

Im **soniXplorer** wird MDS jedoch nur zur Transformation in eine Zwischenrepräsentation verwendet. Die letztliche Platzierung der Stücke und Visualisierung basiert auf einer selbstorganisierenden Merkmalskarte, die auf der Basis der $\vec{x}_i$ gelernt wird. Dies hat u.a. folgende Gründe:

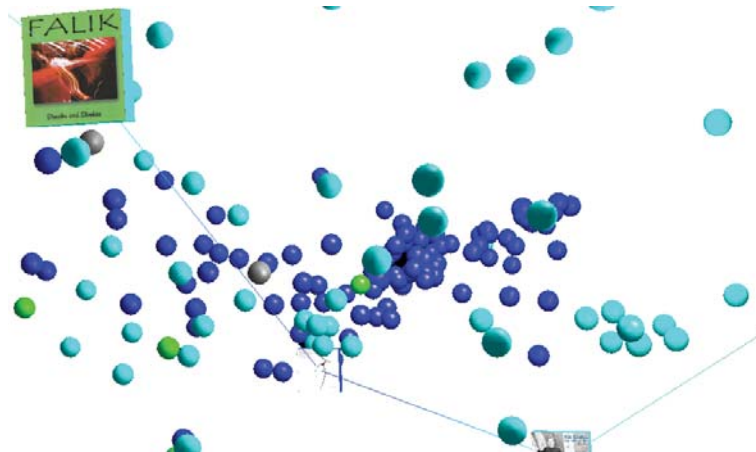


Abbildung 3.8: MDS-basierte dreidimensionale Kollektions-Visualisierung [LE07]

- MDS platziert die betrachteten N Stücke in einem niedrig-dimensionalen Raum. Eine intuitiv erfassbare Strukturierung dieses Raumes, wie sie durch das Höhenprofil einer selbstorganisierenden Karte visualisiert wird, ist nicht verfügbar.
- Zur Trennung von Gruppen unterschiedlicher Stücke positioniert MDS diese in großem Abstand voneinander. Auch SOM-basierte Verfahren sind bemüht, die Ursprungs-Distanzen im zweidimensionalen Zielraum zu reflektieren, sind hier jedoch deutlich flexibler. So lassen sich unterschiedliche Stück-Cluster in deutlich geringerer Entfernung zueinander positionieren. Eine klare optische Trennung ergibt sich aus der Verschiedenheit der Modellvektoren in den Cluster-Regionen. Bei Verwendung einer U-Matrix-Visualisierung wird dies durch die Darstellung einer hohen Bergkette sichtbar. Dies ermöglicht eine kompaktere Präsentation der Kollektion als dies mit MDS-Platzierungen möglich wäre.
- Durch die Anpassung des Höhenprofils einer SOM kann der Benutzer intuitiv Grenzen zwischen Musiktitel-Gruppen ziehen oder auflösen. Dieser Eingriff in die Ähnlichkeitsstruktur ist in MDS-platzierten Visualisierung deutlich schwieriger, da er die Replatzierung einer großen Objektanzahl erfordern würde.

Erste Experimente haben gezeigt, dass bei den untersuchten Kollektionsgrößen von 200 bis 500 Stücken eine Transformation der Unähnlichkeitmatrix in einen 20-dimensionalen Vektorraum ($k = 20$) gute Ergebnisse liefert. Diese Dimensionalität entspricht der Größe des Feature-Raumes, der vom MusicMiner-System zum Aufbau einer emergenten SOM verwendet wird [MUNS05].

3.3 Strukturierung mit selbstorganisierenden Merkmalskarten

Die Entwicklung selbstorganisierender Merkmalskarten geht auf Arbeiten von Kohonen aus dem Jahr 1982 zurück [Koh82]. Sie stellen heute eine populäre Methode des Explorati-

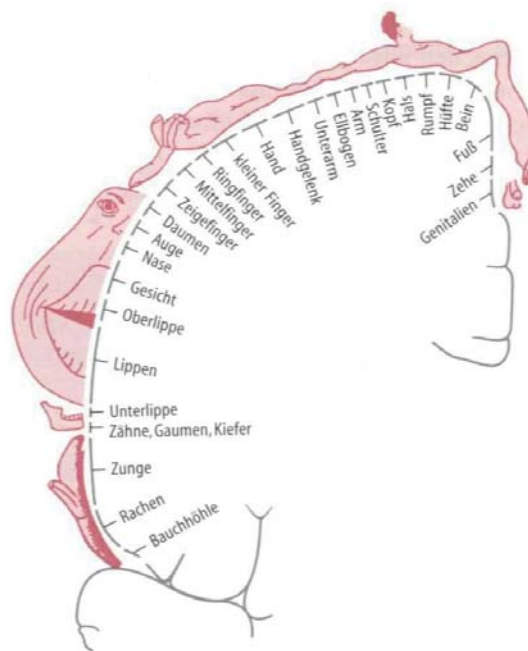


Abbildung 3.9: Entfaltung somato-sensorischer Cortex mit betroffenen Körperregionen (aus [PH01])

ven Data-Mining dar, mit deren Hilfe Datenmengen zu homogenen Gruppen zusammengefasst werden können und die Ähnlichkeitsstruktur von hochdimensionalen Objekten in einer niedrig-dimensionalen Darstellung (meist 2-dimensional) visuell erfassbar wird.

Selbstorganisierende Merkmalskarten sind seit dem intensiv untersucht [Koh01] und in zahlreichen vielfältigen Anwendungskontexten wie Zeichenerkennung [IAD⁺92], Spracherkennung [KS97] oder ingenieurwissenschaftlichen Problemstellungen [KOS⁺96] erfolgreich eingesetzt worden. Eine besondere Bedeutung haben SOMs im Bereich der semantischen Gruppierung und Visualisierung von Dokumentensammlungen erlangt [BSJ00].

Selbstorganisierende Merkmalskarten sind inspiriert durch die aufgabenspezifische Anordnung von Neuronen im menschlichen Hirn. Wie funktionale MRT-Aufnahmen zeigen, ist die Verarbeitung lokaler körperlicher Reize als Folge der lateralen Umfeldhemmung auf abgrenzbare Regionen des Cortex beschränkt. Abbildung 3.9 zeigt eine Entfaltung des somato-sensorischen Rindenfeldes, das für den Tastsinn maßgeblich verantwortlich ist, und stellt seine Bereiche in Bezug zur betroffenen Körperregion.

Es lassen sich zwei Beobachtungen treffen: Zum einen ist der Zusammenhang zwischen dem auslösenden Körperteil und dem Ort der Verarbeitung im Hirn nicht zufällig. Vielmehr sind die Teile des Cortex, die die Tastwahrnehmung benachbarter Körperregionen realisieren, ebenfalls benachbart. Außerdem erkennt man, dass besonders sensiblen Körperteilen wie Hand und Daumen ein für ihre Größe unverhältnismäßig langer Abschnitt der Hirnrinde zugeordnet ist. Diese beiden Charakteristika sind zentrale Eigenschaften der Konstruktion selbstorganisierender Karten: Aufgabe eines SOM-Trainings ist die Generierung einer Abbildung, die hochdimensionale Objekte Topologie-erhaltend, d.h. unter größtmöglicher Wahrung ihrer wechselseitigen Ähnlichkeiten im Ursprungsraum, auf eine niedrigdimensionale Repräsentation abbildet.

Regionen im hochdimensionalen Raum, die eine hohe Objekt-Dichte aufweisen, nehmen im Zielraum ferner eine überproportional große Region ein. Dieser „variable Abbildungsmaßstab“ ermöglicht einen detaillierteren Blick auf Objekte, die ansonsten kaum differenzierbar auf ähnlichen Positionen platziert würden.

Nachbarschaftserhaltung und Dichten-adaptive Abbildung machen selbstorganisierende Merkmalskarten damit zu einem wichtigen Werkzeug der Datenanalyse. Ihre nicht-lineare Dimensionsreduktion hilft bei der Aufdeckung von Strukturen und der Identifikation von Clustern, die sich in den Ursprungsdaten schwer erkennen lassen.

Die Verfügbarkeit intuitiver Karten-Visualisierungen macht sie ferner bedeutsam für den Aufbau einer immersiven Explorationsumgebung.

3.3.1 Formalisierung

Es sei $\{\vec{x}_i\}$ eine Menge von N l -dimensionalen Eingabevektoren, die sich paarweise durch ihre Euklidische Distanz $\|x_{i_1} - x_{i_2}\|$ vergleichen lassen.

Eine selbstorganisierende Merkmalskarte besteht aus M Zellen oder Neuronen $\{y_j\}$, denen jeweils ein Modellvektor (auch Gewichtsvektor genannt) $\vec{m}_j$ aus dem Ursprungsraum zugeordnet ist ($\vec{m}_j \in \mathbb{R}^l$).

Die Zellen $\{y_j\}$ sind in einem k -dimensionalen Raum (meist ist $k = 1$ oder $k = 2$) platziert. Verschiedene Topologie-Alternativen werden in Abschnitt 3.3.3 diskutiert. Sie finden ihren Ausdruck in der Distanz $d_k(j_1, j_2)$, die den Abstand zwischen den Neuronen y_{j_1} und y_{j_2} im $\mathbb{R}^k$ beschreibt.

Diese Distanz wird in einen Nachbarschaftswert h_{j_1, j_2} transformiert, zumeist mit Hilfe einer Gauss-Funktion:

$$h_{j_1, j_2}^{Gauss} = \exp \left(- \left(\frac{d_k(j_1, j_2)}{\sigma} \right)^2 \right),$$

wobei σ die Größe des betrachteten Nachbarschaftsradius beschreibt.

Alternativ kommen zur Umrechnung Cosinus- oder Kegel-Funktion zum Einsatz:

$$h_{j_1, j_2}^{Cosinus} = \begin{cases} \cos\left(\frac{d_k(j_1, j_2)}{\sigma}\right) & \text{für } d_k(j_1, j_2) < \sigma \\ 0 & \text{sonst} \end{cases}$$

$$h_{j_1, j_2}^{Cone} = \begin{cases} 1 - \frac{d_k(j_1, j_2)}{\sigma} & \text{für } d_k(j_1, j_2) < \sigma \\ 0 & \text{sonst} \end{cases}$$

Im Gegensatz zur Kegel-Funktion ordnen Gauss- und Cosinus-Funktion Objekten mit geringer Distanz einen überproportional großen Nachbarschaftswert zu und bewerten enge Nachbarschaft damit höher. **soniXplorer** folgt den Arbeiten von Pampalk [PRM02] und Mörchen [MUNS05] zur Anwendung von SOMs im MIR-Kontext und implementiert die Gauss-Funktion.

Am Ende des im folgenden Abschnitt beschriebenen Lernprozesses steht eine Konfiguration gefundener Modellvektoren $\{\vec{m}_j\}$. Anhand dieser Konfiguration wird ein Eingabeobjekt mit dem Datenvektor $\vec{x}_i$ dem Neuron $bm_i \in \{y_j\}$ zugeordnet, dessen Modellvektor den geringsten Abstand zu $\vec{x}_i$ aufweist. Diese auch als *Best Matching Unit* für $\vec{x}_i$ bezeichnete Zelle wird wie folgt ermittelt:

$$bm_i = \operatorname{argmin}_{j'} \|\vec{x}_i - \vec{m}_{j'}\|$$

3.3.2 Lernverfahren

Ziel der Konstruktion einer selbstorganisierenden Karte ist eine Topologie-erhaltende Dimensionsreduktion der Eingabeobjekte. Dies bedeutet, dass für zwei Datenpunkte $\vec{x}_{i_1}$ und $\vec{x}_{i_2}$ die niedrig-dimensionale Distanz d_k der zugeordneten Neuronen möglichst genau den Abstand im Ursprungsraum $\|\vec{x}_{i_1} - \vec{x}_{i_2}\|$ approximiert:

$$\|\vec{x}_{i_1} - \vec{x}_{i_2}\| \approx d_k(bm_{i_1}, bm_{i_2}) \quad \forall i_1, i_2$$

Hierzu wird die Karte wiederholt mit Eingabeobjekten konfrontiert. Nach der Bestimmung der Best Matching Unit wird der Modellvektor dieser Zelle in Richtung des Datenvektors verschoben. Diese Anpassung dehnt sich (in schwächerem Maße) auf die benachbarten Neuronen aus und erhöht somit die Wahrscheinlichkeit einer folgenden Positionierung ähnlicher Datenobjekte in der Umgebung.

Reduziert man Nachbarschaftsradius und Anpassungsrate (auch als Lernrate bezeichnet) mit der Anzahl durchgeführter Iterationen, konvergieren die Gewichtsvektoren der Neuronen zu einer Konfiguration, die die hochdimensionale Ähnlichkeitsstruktur der $\{\vec{x}_i\}$ in einem k -dimensionalen Zellen-Netz approximiert.

Dieser Anpassungsvorgang lässt sich wie folgt formalisieren:

Sei $\alpha(t)$ eine monoton fallende Lernrate zum Zeitpunkt t . Es hat sich als günstig erwiesen, nach einer Grob-Anpassung mittels hohem α die Lernrate schnell zu senken und zu einer Fein-Abstimmung der Modellvektoren überzugehen. Dieser Verlauf wird z.B. durch einen exponentiellen Zusammenhang zwischen α und t realisiert:

$$\alpha(t) = \alpha_0 \cdot e^{-\frac{t}{t_{max}}},$$

wobei α_0 die Lernrate zu Beginn des Trainings darstellt und t_{max} die Anzahl an Iterationsschritten beschreibt.

In analoger Weise reduziert man den Radius σ in der Nachbarschaftsfunktion h_{j_1, j_2} mit fortschreitender Zeit, um die Anpassung der Modellvektoren zum Ende hin auf lokale Feinkorrekturen zu beschränken.

Mit diesen Definitionen lässt sich ein erster SOM-Trainings-Algorithmus formulieren:

In jeder von insgesamt t_{max} Iterationen werden alle N Eingabevektoren $\{\vec{x}_i\}$ in zufälliger Reihenfolge nacheinander betrachtet. Der Datenpunkt $\vec{x}_i$ bewirkt dabei eine Aktualisierung aller M Modellvektoren $\{\vec{m}_j\}$ nach folgender Vorschrift:

$$\vec{m}_j(t+1) := \vec{m}_j(t) + \alpha(t) \cdot h_{j, bm_i}(t) \cdot (\vec{x}_i - \vec{m}_j(t))$$

Die Verarbeitung eines einzelnen betrachteten Eingabevektors lässt sich in zwei Schritte zerlegen: Die Bestimmung der Best Matching Unit und die Anpassung der Modellvektoren dieser Zelle und ihrer Umgebung.

Im Gegensatz zu obigem *inkrementellen* SOM-Trainingsalgorithmus berechnet das sog. *Batch*-Verfahren zunächst die Best Matching Units *aller* Eingabevektoren und gleicht anschließend die Modellvektoren in einem Schritt an [Koh92]:

$$\vec{m}_j^{batch}(t+1) := \frac{\sum_i h_{j, bm_i}(t) \cdot \vec{x}_i}{\sum_{i'} h_{j, bm_{i'}}(t)}$$

Modellvektoren $\vec{m}_j$, bei denen in obiger Formel der Nenner 0 wird, bleiben konstant.

Diese Variante benötigt zum einen keine Parametrisierung einer Lernrate und lässt sich ferner durch die Zusammenfassung von Eingabevektoren in Voronoi-Mengen effizienter implementieren:

Hierzu beschreibe V_j die Menge aller Eingabevektoren $\vec{x}_i$, deren Best Matching Unit die Zelle y_j ist:

$$V_j = \{\vec{x}_i | bm_i = j\}$$

Alle zu einer solchen Menge zusammengefassten Vektoren lassen sich zu einem Zellen-spezifischen l -dimensionalen Repräsentanten aufsummieren:

$$\vec{s}_k = \sum_{\vec{x} \in V_k} \vec{x}$$

Damit lässt sich die Aktualisierungsregel kompakter formulieren:

$$\vec{m}_j^{batch_2}(t+1) := \frac{\sum_k h_{j,k}(t) \cdot \vec{s}_k}{\sum_{k'} h_{j,k'}(t) |V_{k'}|}$$

An die Stelle einer Summation über N Eingabevektoren tritt hier die Summation über M Voronoi-Mengen.

Sowohl Batch-Algorithmus als auch inkrementelles Verfahren haben eine Laufzeitkomplexität von $\mathcal{O}(N \cdot M)$. Ist die Anzahl an Datenpunkten N kleiner als die Anzahl der SOM-Zellen M , macht eine Verwendung des Batch-Verfahrens Sinn. Pampalk et al. verwenden diese Variante zum Aufbau ihrer Island-of-Music-Visualisierungen [PRM02]. Aufgrund der geringen Größe der selbstorganisierenden Karte kommt die SOM-Anpassung einer Gruppierung der Stücke in wenige (zueinander in Beziehung gesetzte) Cluster gleich.

Um eine detailliertere Visualisierung der Ähnlichkeitsstruktur zu ermöglichen, werden im **soniXplorer** dagegen die in Abschnitt 3.1.4 genannten *emergenten* selbstorganisierenden Karten eingesetzt, deren Neuronenzahl deutlich höher ist. Wie bei Mörchen et al. kommt daher der inkrementelle Algorithmus zum Einsatz [MUNS05].

Vor einer sukzessiven Anpassung der Modellvektoren müssen diese geeignet initialisiert werden. Ein erster Ansatz sieht die zufällige Belegung der Gewichtsvektoren vor und überlässt die Suche nach geeigneten $\vec{m}_j$ somit vollständig der iterierten Anwendung der Aktualisierungsregel.

Informierter lassen sich die Modellvektoren mit den Prototypen eines vorgeschalteten k-Means-Clustering-Verfahrens initialisieren. Es bleibt jedoch offen, wie diese Repräsentanten zueinander beziehungsloser Cluster auf der Karte zu verteilen sind.

Im **soniXplorer** kommt eine dritte Alternative zum Einsatz: Zunächst werden mit Hilfe einer Hauptkomponentenanalyse die orthogonalen Richtungen $\vec{a}$ und $\vec{b}$ ermittelt, die den größten Anteil der Varianz in $\{\vec{x}_i\}$ erklären. Unter Beachtung des Wertebereichs der Eingabevektoren werden die Modellvektoren der Neuronen anschließend mit einer Linearkombination aus $\vec{a}$ und $\vec{b}$ initialisiert, die die Position der Zelle auf der Karte widerspiegelt.

3.3.3 Karten-Topologie

Die Anordnung der Zellen auf einer selbstorganisierenden Karte wird durch die Nachbarschaftsfunktion bestimmt. Kohonen beschreibt zwei alternative Topologien, die in Abbildung 3.10 illustriert sind [Koh01].

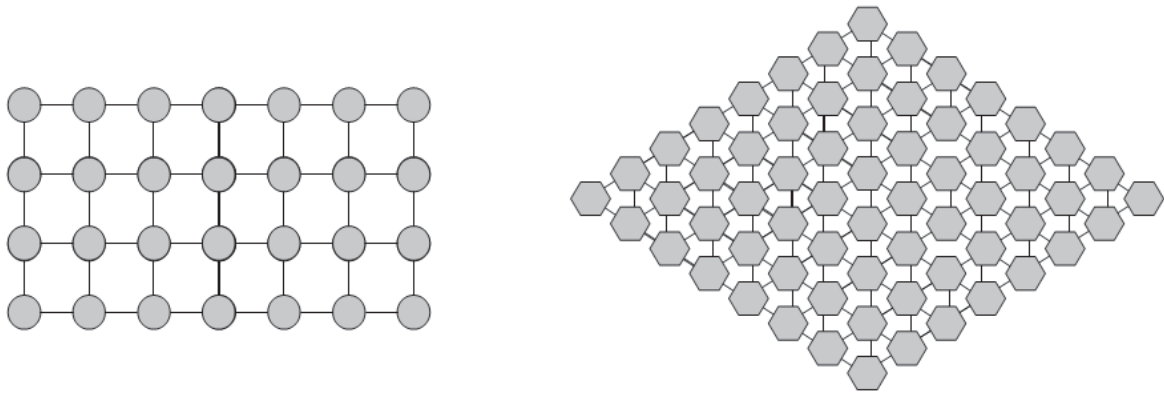


Abbildung 3.10: Rechteckige und hexagonale Netzwerktopologie

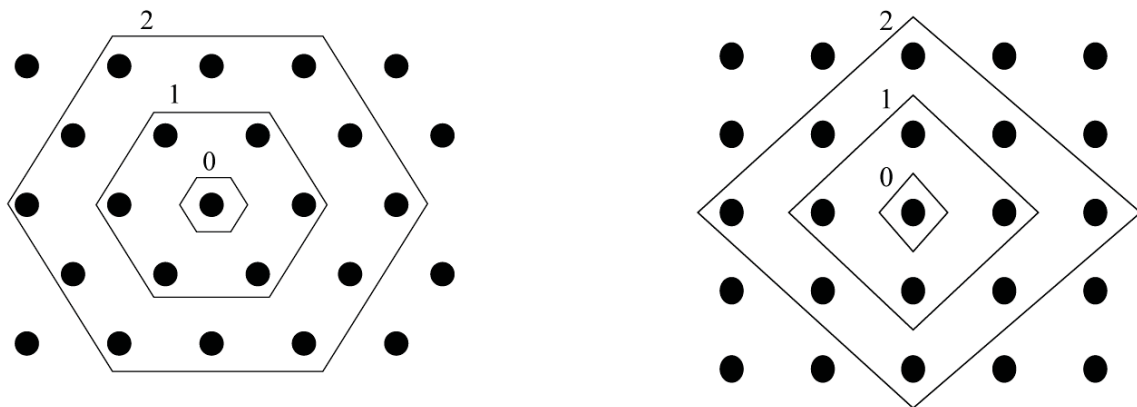


Abbildung 3.11: Regionen gleichen Abstandes vom zentralen Neuron bei hexagonaler und rechteckförmiger Zellen-Topologie

Wählt man eine regelmäßige gitterförmige Anordnung mit rechteckigen Zellen, so sind jedem Neuron vier direkte Nachbarn zugeordnet. Eine gleichmäßigere Verteilung der Zellen auf der Karte erreicht man durch eine hexagonale Struktur, bei der jedes Neuron von sechs anderen Zellen direkt umgeben ist (vergl. auch [SN09]).

Diese Topologie bietet im Allgemeinen Vorteile bei der Visualisierung. Zur Illustration stellt Abbildung 3.11 die Regionen gleichen Abstandes von einem zentralen Neuron dar. Wie sich erkennen lässt, approximieren die Sechsecke die ideale Form einer kreisförmigen Umgebung genauer als die Rhomben bei einer gleichmäßig rechteckförmigen Gitterstruktur.

Diese Diskretisierungsproblematik ist im *soniXplorer* jedoch von geringerer Bedeutung, da zum einen die Neuronen-Dichte bei Verwendung emergenter SOMs relativ groß ist und zum anderen zur Generierung eines durchgängigen Höhenprofils ohnehin zwischen den Einzel-Zellen-Werten interpoliert wird. Daher kommt (wie auch beim Musicminer [MUNS05]) ein rechteckförmiges Gitter zum Einsatz.

Ordnet man die Zellen wie in Abbildung 3.10 dargestellt auf einer begrenzten planaren Fläche

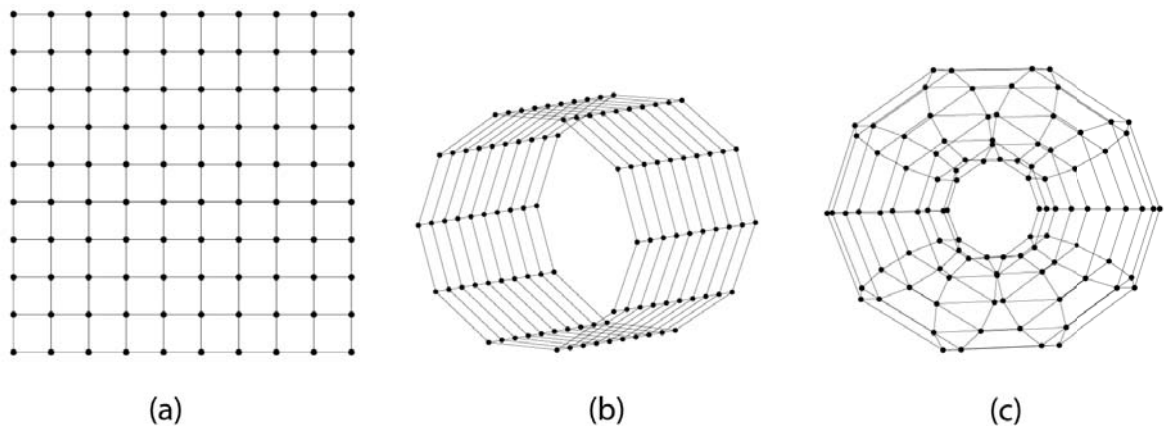


Abbildung 3.12: Visualisierung geschlossener Karten-Topologien: (a) offenes planares Gitter, (b) halb-geschlossene Struktur: Cylinder, (c) vollständig geschlossener Körper: Ring (Toroid)

an, so besitzen randständige Neuronen natürlicherweise weniger benachbarter Zellen. Damit ist auch die Anzahl der Zellen reduziert, in deren Nachbarschaftsmenge sie sich befinden. Dies führt im SOM-Training zu einer geringeren Anpassungswahrscheinlichkeit der außenstehenden Neuronen und letztlich zu lokal stabileren Modellvektor-Zuweisungen. Anschaulich lässt sich dies als ein einseitiges „Drücken“ von Objekten in Randbereiche veranschaulichen, dem keine inverse Kraft entgegen tritt. Die daraus resultierende Häufung von Objektplatzierungen an der Kartengrenze wird als *Randeffekt* bezeichnet [Ult03a].

Diese Problem lässt sich umgehen, indem man randständige Neuronen mit den Zellen der gegenüberliegenden Seite zu einer geschlossenen Karten-Topologie verbindet. Abbildung 3.12 illustriert die entstehenden (dreidimensionalen) Strukturen bei Verknüpfung des linken und rechten Randes (Cylinder) und bei einer zusätzlichen Verbindung der oberen und unteren Grenzen (Toroid).

Um zu einer zweidimensionalen Visualisierung zurückzugelangen, müssen diese dreidimensionalen Objekte aufgetrennt und ihre Hülle planar dargestellt werden. Während des SOM-Trainings benachbarte Neuronen werden durch diese Trennung auf gegenüberliegenden randständigen (und damit weit voneinander entfernten Positionen) platziert. Ultsch beschreibt ein heuristisches Verfahren, um diese Separierung möglichst weitgehend entlang von Cluster-Grenzen durchzuführen, die ohnehin eine Abgrenzung von Datenbereichen andeuten [Ult03a]. Dies führt zu planaren Karten mit nicht-rechteckigem Umriss (vergl. Abbildung 3.4). Wie dort zu erkennen ist, kann das Trennung-Verfahren von Ultsch die Problematik jedoch nicht vollständig lösen: Beispielhaft sei die räumliche Trennung des eigentlich zu den anderen Rap-Stücken benachbarten türkisfarbenen Punktes in der unteren Mitte der Karte angeführt.

Wie sich in ersten Experimenten mit nicht-geschlossenen Topologien herausstellte, ist der beschriebene Randeffekt bei geeigneter Parametrisierung von Lernrate und zeitabhängigem Nachbarschaftsradius beherrschbar. Ferner können bereits kleine Änderungen der Modellvektoren wie sie z.B. durch eine Anpassung durch oder an den Benutzer entstehen (siehe Kapitel 4) zu großen Veränderungen der Hüllen-Darstellung führen. Da eine massive Modifikation der Visualisierung bei nur kleinen Aktualisierungen dem Anwender die Orientierung in der Explorationslandschaft erschwert und aufgrund ihrer Unnatürlichkeit die Immersion mit der

virtuellen Welt zerstört, wurde im soniXplorer auf die Verwendung geschlossener Topologien verzichtet.

3.3.4 Höhenprofil-Berechnung

Selbstorganisierende Merkmalskarten ermöglichen die Abbildung hochdimensionaler Eingabevektoren auf niedrigdimensionale Koordinaten und können daher als Dimensionsreduktionsverfahren betrachtet werden. Über diese Projektion der Objekte hinaus trägt die generierte Karte kraft ihrer Modellvektoren jedoch weitere Information, die zur Veranschaulichung der Ähnlichkeitsstruktur der Kollektion verwendet werden kann.

In der Literatur werden verschiedene Ansätze diskutiert, die den einzelnen Zellen einen „Intensitätswert“ zuweisen, der eine anschließende Visualisierung der Karte erlaubt.

Diese Verfahren lassen sich in zwei Gruppen klassifizieren:

1. Dichte-basierte Verfahren

Hintergrund dieser Ansätze ist die Interpretation der Eingabevektoren als Stichproben einer Grundgesamtheit. Auf der Grundlage der Anzahl von beobachteten Datenobjekten, die einer SOM-Zelle zugeordnet wurden, lassen sich mit dieser Auffassung Schätzverfahren für die tatsächliche Dichte formulieren.

Utsch stellt in [Ult03b] eine *Pareto Density Estimation* genannte Prozedur vor, die die Datenobjekt-Dichte eines Neurons aus der Zählung beobachteter Zuweisungen im Umkreis der Zelle errechnet. Je größer der Radius dieses Kreises ist, umso mehr Daten liegen für eine fundiertere Schätzung vor. Mit wachsender Umgebungsgröße nimmt auf der anderen Seite aber die örtliche Auflösung der Schätzung ab.

Utsch wendet ein informations-theoretisches Maß zu Bestimmung eines Kompromisses zwischen diesen konträren Forderungen an. Seine *Pareto-Dichten-Schätzung* setzt den Radius des Schätz-Kreises auf einen konstanten Wert, der durchschnittlich 20% der Datenmenge umfasst [Ult03b]. Die Zusammenstellung der so geschätzten Dichten für jedes Neuron wird als *P-Matrix* bezeichnet [Ult03a].

Während die Kern-basierte Dichte-Schätzung (kernel-based density estimation) von Utsch die Berechnungsgrundlage auf benachbarte Neuronen ausweitet, betrachten Pampalk et al. Eingabevektoren, die auf Zellen mit ähnlichem Modellvektor projiziert wurden [PRM02]:

Bei vorgegebenem n wird hierzu für jedes Objekt der Zähler seiner Best Matching Unit um n erhöht. Der Zähler der Zelle, die am zweit-ähnlichsten zum Eingabevektor ist, wird um $n - 1$ inkrementiert, usw.

An die Stelle einer starren Zuordnung von Eingabeobjekt zu SOM-Zelle tritt somit eine graduelle Verteilung auf Neuronen, die sich nach dem Rang der Ähnlichkeit zum Eingabevektor richtet. Da hieraus eine Art Glättung der Anzahl zugewiesener Objekte pro Neuron resultiert, ist dieses Verfahren unter der Bezeichnung *Smoothed Data Histograms* bekannt.

2. Distanz-basierte Verfahren

Cluster lassen sich auf der Karte nicht nur durch eine erhöhte Dichte zugewiesener

Datenobjekte identifizieren, sondern äußern sich auch in deutlichen Modellvektor-Veränderungen an Cluster-Grenzen. Die Visualisierung der Distanzen zwischen den Gewichtsvektoren benachbarter Zellen kann daher zur Beschreibung der Kartenstruktur herangezogen werden.

Die auf Ultsch zurückgehende populäre U-Matrix führt für jede Zelle den durchschnittlichen Abstand ihres Modellvektors zu den Gewichtsvektoren der unmittelbaren Nachbarn [Ult92] auf.

Beschreibt man die Menge der Nachbarn einer Zelle y_j mit U_j , lässt sich der U-Matrix-Wert für das Neuron y_j wie folgt berechnen:

$$u_j = \frac{1}{|U_j|} \cdot \sum_{y_{j'} \in U_j} \|\vec{m}_j - \vec{m}_{j'}\|$$

Wie Becks et al. berichten, führt die Visualisierung des minimalen Abstandes v_j zu benachbarten Modellvektoren im Kontext der Analyse von Textsammlungen häufig zu klareren Cluster-Grenzen [BSJ00]:

$$v_j = \min_{y_{j'} \in U_j} (\|\vec{m}_j - \vec{m}_{j'}\|)$$

Obige Unterscheidung in Distanz- und Dichte-basierte Verfahren ist zwar angesichts der verschiedenen Berechnungsansätze begründbar, letztlich sind beide Gruppen aber eng miteinander verbunden, denn die iterative Anpassung der SOM im Lernprozess führt dazu, dass Zellen in Regionen mit hoher Objektdichte in der Regel ähnliche Modellvektoren aufweisen.

In initialen Experimenten erwies sich keine der obigen Alternativen im Kontext einer Musikkollektions-Visualisierung als eindeutig überlegen. Daher folgt der *soniXplorer* der Strategie des MusicMiners [MUNS05] und ordnet einem Neuron y_j seinen U-Matrix-Wert u_j zu. Dieser lässt sich als suggestive Färbung der Zelle auf einer zweidimensionalen Landkarte darstellen oder - wie im folgenden beschrieben - als Höhenwert einer dreidimensionalen Landschaft interpretieren, in der Täler ähnlicher Stück-Cluster durch Bergketten voneinander separiert sind.

3.4 Multimodale Präsentation

Populäre Visualisierungen selbstorganisierender Merkmalskarten wie etwa der Islands-of-Music-Ansatz [PRM02] oder die MusicMiner-Arbeiten [MUNS05] basieren auf einer übersichtartigen zweidimensionalen Darstellung der Musikkollektion. Nach Ähnlichkeit gruppierte Stücke werden als Punkte in einer aus den Modellvektoren abgeleiteten Landschaftsstruktur symbolisiert.

Diese Visualisierung weist zwei Schwächen auf:

- Die akustische Präsentation eines Musiktittels beschränkt sich auf die Möglichkeit des Benutzers, ein Stück bewusst zur Einzel-Wiedergabe auszuwählen. Diese Einschränkung berücksichtigt damit nicht die Fähigkeit des Menschen, einzelne Klänge aus einem

akustischen Gemisch zu isolieren und sich in einer (ggf. virtuellen) Welt aus Schallquellen zu orientieren, die aus verschiedenen Richtungen ertönen („Cocktail-Party-Effekt“ [Aro92, Che53]).

Die Annotation positionierter Stücke mit Titel-, Künstler- oder Genre-Angaben ist stets eine Darstellung von umschreibenden Metadaten und resultiert unweigerlich in einem Medienbruch zwischen dem Klang der Musik und seiner textuellen Repräsentation.

Informativer und intuitiver ist die Wiedergabe der Musik selbst zur Darstellung eines Stückes („Music is described best by music“ [Lü05]).

- Die Farbgebung der Karten-Zellen suggeriert zwar Cluster-Grenzen, verlangt jedoch die Interpretation als Höhenprofil durch den Benutzer. Intuitiver ist eine direkte dreidimensionale Visualisierung der Karte, in der Separierungen zwischen Gruppen ähnlicher Musik als Bergketten klar zu erkennen sind.

Die Konstruktion einer navigierbaren dreidimensionalen Umgebung, in der visuelle Informationen durch die Positionierung simultan erklingender Schallquellen ergänzt werden, ermöglicht die Immersion des Benutzers in diese virtuelle Welt und fördert den Explorationsdrang des Anwenders.

An die Stelle der distanzierten Betrachtung einer zweidimensionalen Karte tritt somit das Erlebnis einer Computerspiel-artigen Landschafts-Simulation, in die der Benutzer eintauchen und sich geleitet von akustischen und visuellen Hilfestellungen frei bewegen kann.

Ziel dieser Arbeit ist die Entwicklung und Evaluierung einer solchen multimodalen Explorationsumgebung.

Diese Entwicklung folgt einem iterativen Prozess: In einem ersten Java-basierten Prototypen wurden Visualisierungsideen erprobt und in einer kleinen Benutzerstudie ausgewertet [Nei07]. Erfahrungen mit diesem System fanden Eingang in die anschließende Implementierung des *soniXplorers*, der als .NET-Anwendung in Microsofts Gaming-Framework XNA eine hinreichend performante Darstellung auch größerer Musikkollektionen erlaubt.

In ähnlicher Weise wurden Auralisierungskonzepte zunächst in kleinen Prototypen entwickelt und getestet [Lü05]. Die Implementierung eines Fokussierungs-Mechanismus, der bei simultaner Wiedergabe möglichst vieler Stücke die Gefahr einer akustischen Reizüberflutung vermeidet, wurde so konzentriert erprobt und schließlich in den *soniXplorer* integriert.

Im folgenden werden Visualisierung und akustische Präsentation einer Musiksammlung, wie sie im *soniXplorer* realisiert sind, vorgestellt.

Zur Illustration dient dabei die in Tabelle A.1 aufgeführte kleine Kollektion von Stücken unterschiedlicher Genres. Abbildung 3.13 zeigt die generierte Positionierung dieser Titel auf einer selbstorganisierenden Merkmalskarte, deren Zellenfärbung sich aus einer Interpretation der U-Matrix-Einträge als Graustufenwerte ergibt.

Deutlich zu erkennen ist die Gruppierung von Stücken gleichen Genres; mitunter trennt die gezeigte SOM Titel eines Genres jedoch in verschiedene Cluster (vergl. z.B. die Separierung der beiden Songwriter-Cluster in der oberen Mitte sowie die Abgrenzung zu den drei Stücken am rechten Rand der Karte).

In der Gesamtheit zeigt Abbildung 3.13 jedoch eine recht gute Korrespondenz zwischen der automatisch erzeugten Gruppen-Bildung und der (zum SOM-Training nicht verwendeten) Genre-Klassifikation.

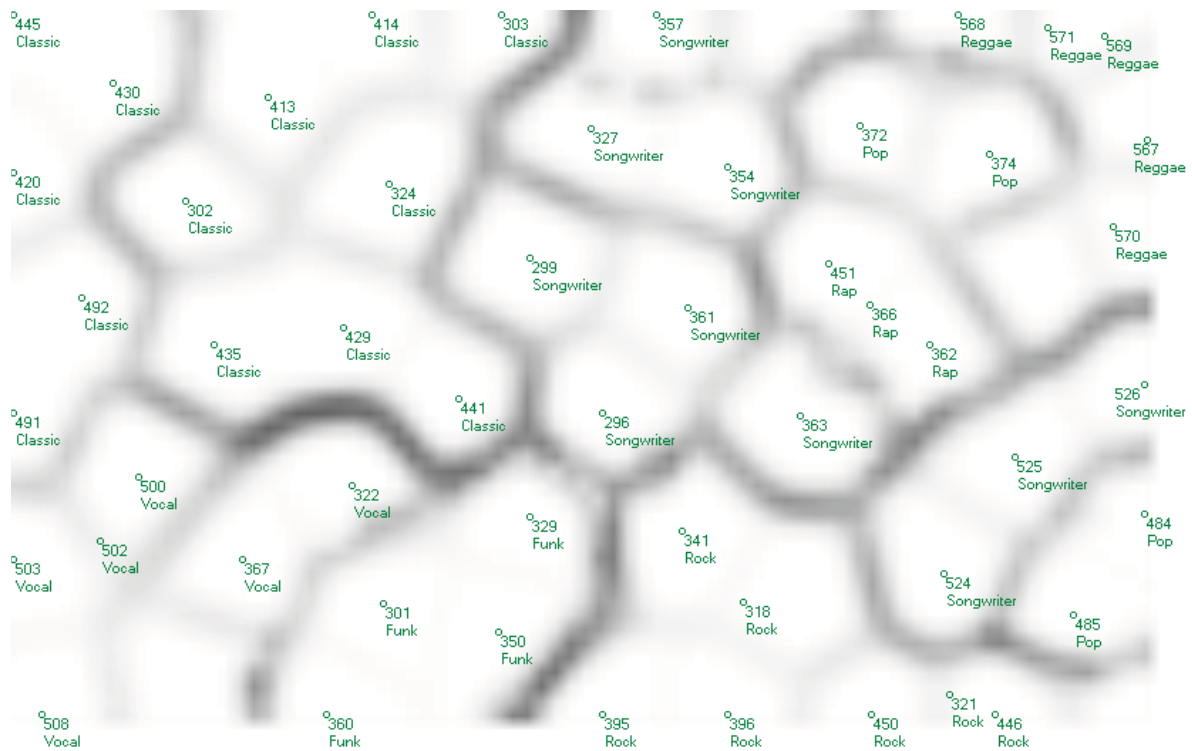


Abbildung 3.13: Darstellung der Demonstrationskollektion auf einer zweidimensionalen selbst-organisierenden Merkmalskarte mit Graustufenfärbung



Abbildung 3.14: Darstellung der Demonstrationskollektion in der soniXplorer-Explorationsumgebung. Blick vom linken Kartenrand in die Mitte der SOM.

3.4.1 Visuelle Darstellung

Abbildung 3.14 vermittelt einen ersten Eindruck der implementierten Explorationsumgebung. Aus einer halb-hohen Position ist der Blick auf eine Landschaft aus Bergen und Tälern zu sehen, in der sich die als Würfel dargestellten Stücke der beschriebenen Demonstrationskollektion befinden. Das Bild stellt die Perspektive eines virtuellen Betrachters dar, der vom linken oberen Rand in die Mitte der Karte schaut. Deutlich zu erkennen ist die hohe Bergkette auf der rechten Seite, die die Klassik-Region von den nicht sichtbaren dahinter liegenden Vokal-Stücken trennt.

Zur Abgrenzung des Terrains ist die Landschaft von einem Meer umgeben. Ferner erzeugt eine um den Beobachter positioniertes Halbkugel („SkyDome“) mit überlagerten Texturen weißer Gauss-verteilter Pixel-Haufen die Illusion eines Himmels mit langsam vorüberziehenden Wolken. Die Beleuchtung der Szene wird durch eine Phong-Shading-Modell realisiert [Pho75]: Eine Kombination von Umgebungslicht und directionalem Licht, das zu einer Abschattung der Lichtquellen-abgewandten Bergseiten führt, verstärkt den dreidimensionalen Eindruck der Landschaft.

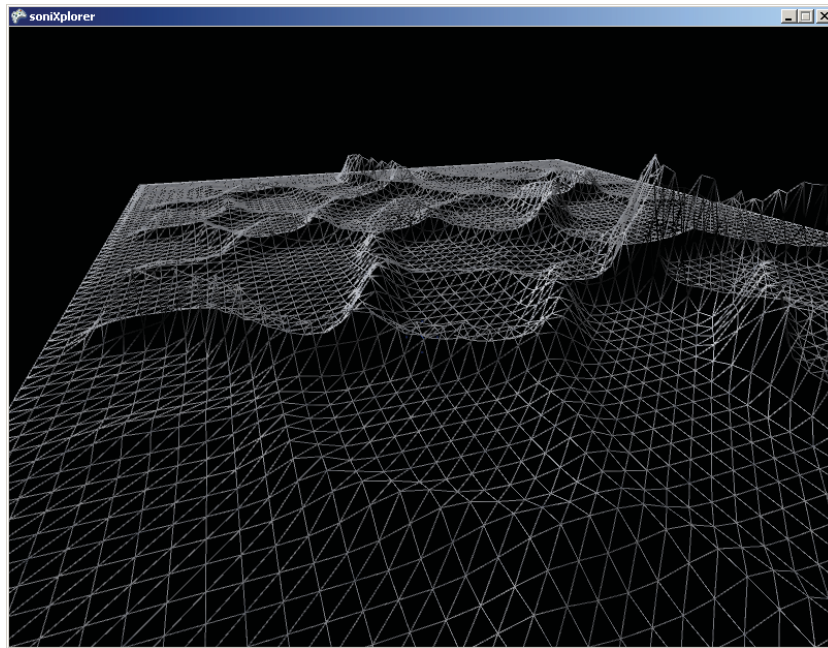


Abbildung 3.15: Drahtmodell-Darstellung ohne Höhenprofil-Interpolation

Die folgenden Abschnitte beschreiben die zentralen Komponenten der soniXplorer -Visualisierung detaillierter.

Terrain

Zur Darstellung einer Landschaft aus Bergen und Tälern wird ein Netz aus verbundenen Dreiecken verwendet. Die räumliche Positionierung der Eckpunkte ergibt sich aus dem Höhenprofil, das durch die U-Matrix der selbstorganisierenden Karte bestimmt ist. Abbildung 3.15 zeigt das Drahtmodell eines so generierten Terrains.

Trotz der vergleichsweise hohen Neuronen-Dichte emergenter SOMs zeigt sich, dass die Auflösung des Dreiecks-Netzes für eine authentische Landschaftsdarstellung zu grob ist. Daher ermöglicht der soniXplorer die Interpolation zusätzlicher Zwischen-Gitterpunkte. Um einen möglichst ruhigen und gleichmäßigen Höhenverlauf zu erzeugen, wird auf eine zweidimensionale Catmull-Rom-Interpolation zurückgegriffen [CR74].

Mit einer Verfeinerung der Höhenprofil-Auflösung um den Faktor r vervielfacht sich die Anzahl der Gitterpunkte um r^2 und die Anzahl der benötigten Dreiecke im Netz um $2 \cdot r^2$. Daher muss ein Kompromiss zwischen Darstellungsqualität und Szenenkomplexität, die die Rendering-Geschwindigkeit und damit die Reaktivität der Umgebung beeinflusst, gewählt werden. Aus diesem Grund ist die Auflösungsverfeinerung r im soniXplorer durch den Benutzer konfigurierbar. Bei aktuellen PCs und einer Kartengröße von $80 \cdot 50$ Zellen ergaben sich für $r = 3$ keine spürbaren Geschwindigkeitseinbußen. Das in Abbildung 3.16 gezeigte Drahtmodell für ein so interpoliertes Höhenprofil veranschaulicht die Verbesserung der Darstellungsqualität des Terrains.

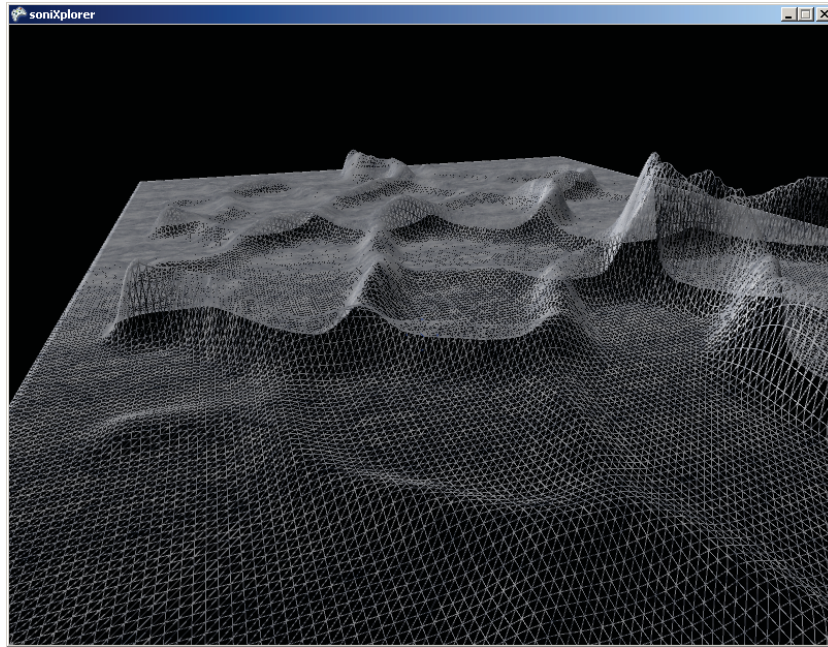


Abbildung 3.16: Drahtmodell-Darstellung mit Höhenprofil-Interpolation

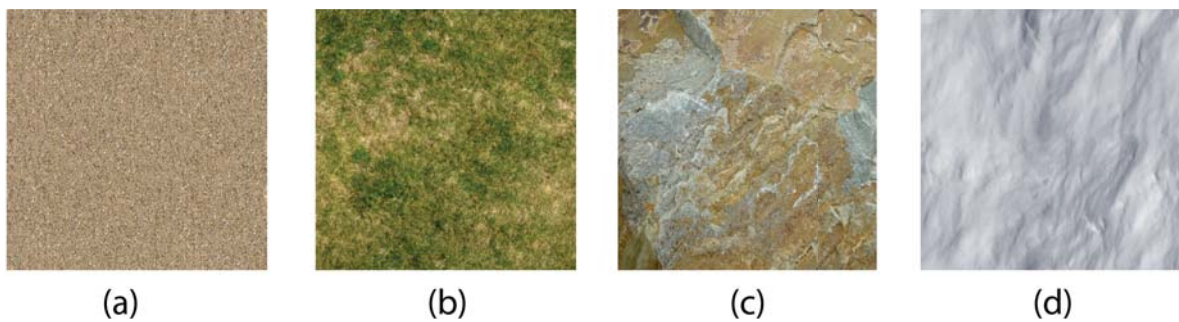


Abbildung 3.17: Terrain-Texturen

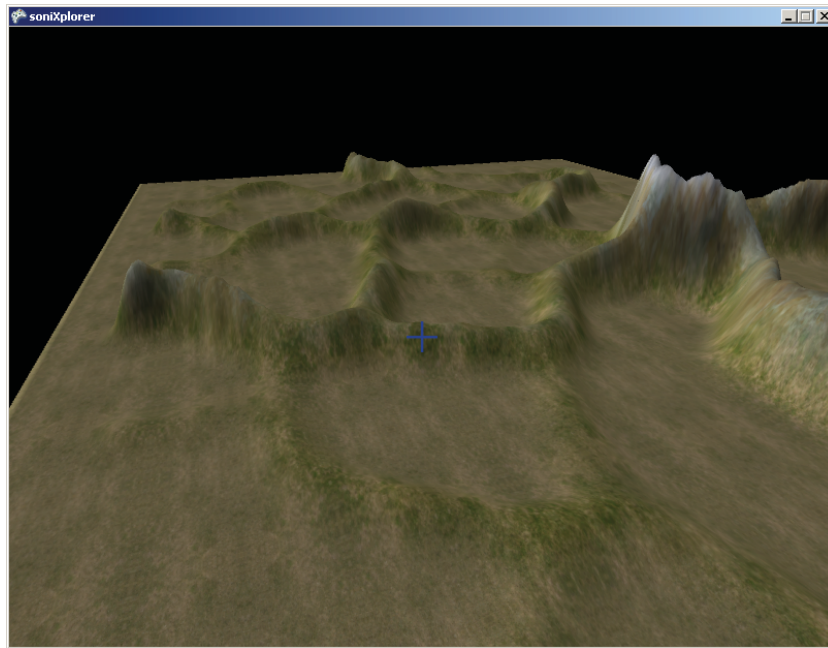


Abbildung 3.18: Texturiertes Terrain

Um die Illusion von Bergen und Tälern zu verstärken, wird dieses Drahtmodell mit einer höhenabhängigen Kombination von Einzeltexturen bedeckt, die in Abbildung 3.17 dargestellt sind. Mit steigender Höhe wechselt die Oberfläche von Sand (a) über Gras (b) und Berggestein (c) bis zu schneebedeckten Berg-Kuppen (d). Ein eigens entwickelter Pixel-Shader verlagert das Rendering des texturierten Terrains in die Grafikkarte und ermöglicht somit eine effiziente Darstellung. Abbildung 3.18 zeigt das Resultat dieses Rendering-Prozesses.

Stücke

Die Stücke der zu visualisierenden Musikkollektion werden im **soniXplorer** standardmäßig als kurz über dem Terrain schwebende schräg gestellte Würfel dargestellt (vergl. Abbildung 3.14). Im ersten Prototyp der Explorationsumgebung wurden Kugeln als Stück-Repräsentanten verwendet. Es zeigte sich jedoch, dass die Komplexität eines Polygon-Netzes zur Approximation einer Kugel die Renderingperformanz bei größerer Titellanzahl deutlich senkt. Ein weiterer Vorteil des Würfels besteht in der Möglichkeit, seine planaren Seiten mit gut lesbaren Texturen zu bedecken. So verwendet der **soniXplorer** Albencover, die nach erfolgreicher automatischer Stück-Identifikation (vergl. Abschnitt 2.8.2) durch Amazons Web Services zur Verfügung gestellt werden, als Würfel-Hülle zur visuellen Beschreibung von Stücken. Alternativ wird aus Titel und Künstler des Songs eine Texture gebildet.

Aufgrund der Schrägstellung der Würfel lässt sich (im Gegensatz zur Kugel) ferner sehr schnell entscheiden, ob das Objekt um die eigene Achse rotiert. Diese Eigenschaft lässt sich damit zur Darstellung eines binären Stück-Attributes nutzen. Im **soniXplorer** dient sie zur Kennzeichnung der Titel, die aktuell akustisch wiedergegeben werden (vergl. Abschnitt 3.4.2).

Stücke, deren Platzierung in der Musiklandschaft dem Anwender nicht angebracht erscheint, können zur Neupositionierung in der nächsten SOM-Iteration, die möglicherweise auf einem

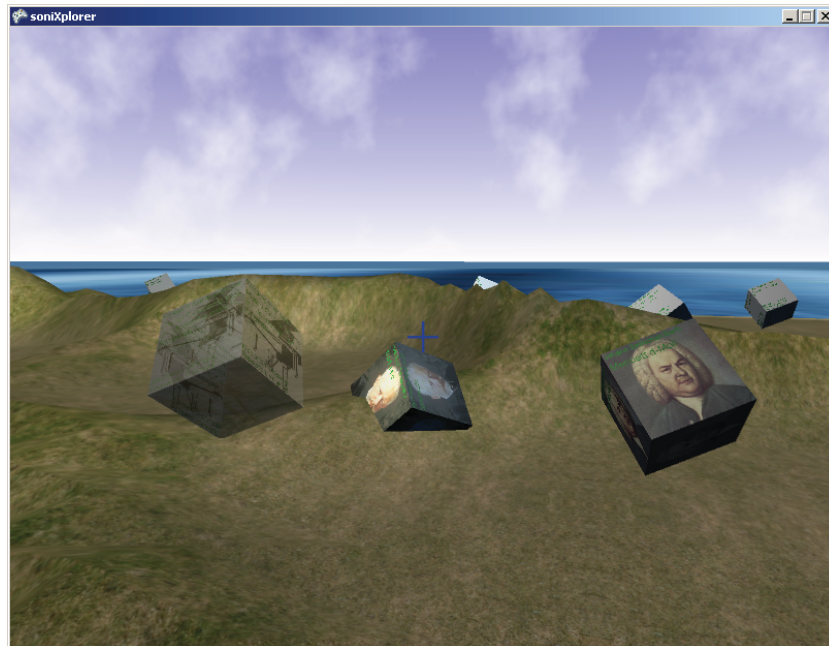


Abbildung 3.19: Visualisierung von Stück-Eigenschaften durch Transparenz, Arretierung und Blinken

aktualisierten Ähnlichkeitsmaß-Mix basiert (vergl. Kapitel 4) freigesetzt werden. Diese Lösung wird im soniXplorer durch die Erhöhung der Transparenz des Stück-Symbols und die daraus resultierende blasse Darstellung suggeriert.

Wie im folgenden Abschnitt 3.4.2 detaillierter beschrieben, wählt die Explorationsumgebung abhängig vom aktuellen Standpunkt des Benutzers automatisch eine Menge simultan wiedergegebener Stücke aus. In diese Selektion kann der Anwender eingreifen, indem er bewusst bestimmte Titel von der Wiedergabe ausschließt. Dies wird durch die partielle Absenkung des Würfels in das Terrain visualisiert. Die hierdurch angedeutete Arretierung im Boden verhindert eine Drehbewegung des Würfels, die wie beschrieben für die Wiedergabe des Stückes steht. In ähnlicher Weise kann der Benutzer Titel permanent zur Wiedergabeliste hinzufügen. Solche „daueraktiven“ Stücke werden als blinkende Würfel dargestellt.

Abbildung 3.19 zeigt diese Eigenschaftsvisualisierungen (Transparenz, Arretierung und Blinken) am Beispiel von drei Stücken im Vordergrund.

Die Platzierung eines Stückes auf dem Terrain basiert auf der Position der Best-Matching-Zelle. Um eine Konzentration der Würfel auf den Gitterpunkten der Karte zu vermeiden, interpoliert soniXplorer bilinear zwischen den Modell-Vektoren der benachbarten Neuronen und wählt in diesem Feld die Position mit geringster Modell-Vektor-Distanz zum Titel aus.

3.4.2 Akustische Darstellung

Die visuelle Präsentation von Stücken als beschriftete Würfel ermöglicht dem Benutzer zwar eine erste Orientierung in der virtuellen Umgebung; informierter kann er die Musiksammlung

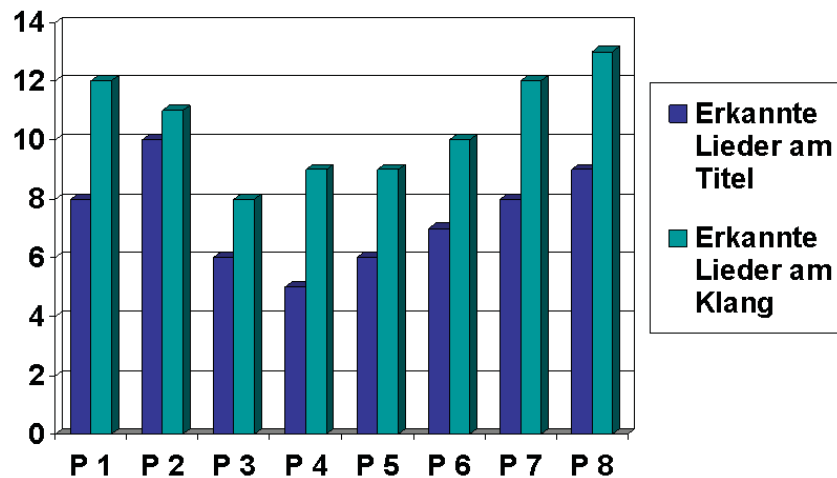


Abbildung 3.20: Gegenüberstellung der Anzahl erkannter Lieder am Titel vs. erkannter Lieder am Klang [Nei07]

jedoch explorieren, wenn diese Symbole gleichzeitig als Schallquellen fungieren, die den dargestellten Titel auch akustisch wiedergeben. Diese multimodale Präsentation vereinfacht die Immersion des Anwenders in die virtuelle Welt und stimuliert seinen „Entdeckergeist“.

Die Bedeutung der akustischen Präsentation akustischen Materials wurde im Rahmen der Evaluation des ersten *soniXplorer*-Prototypen herausgearbeitet: Neis konfrontiert hierzu acht Probanden mit einer Liste von 20 Musikstücken und ermittelt den Anteil dieser Kollektion, der den Testpersonen rein aufgrund der Stück-Metadaten vertraut ist. Anschließend haben die Probanden die Möglichkeit, sich die Stücke anzuhören. Wie Abbildung 3.20 zeigt, erkennen alle Evaluationsteilnehmer durch diese akustische Information weitere vorher nicht genannte Stücke wieder. Dieses Ergebnis unterstreicht die Forderung nach einer Medienbruch-freien Präsentation von Musiktiteln für eine intuitive und effiziente Exploration.

Positionierung der Schallquellen im virtuellen Raum

Die Position der Schallquelle, die ein Stück omnidirektional wiedergibt, fällt mit dem Zentrum seiner Würfelkoordinaten zusammen. Abhängig von diesen Koordinaten und der aktuellen Position des Benutzers muss eine Verteilung von Signalanteilen auf die verfügbaren Wiedergabekanäle erreicht werden, die die Illusion eines Klangesursprungs aus der entsprechenden Richtung erzeugt.

Im Kontext einer öffentlichen Installation in einem Museum verwenden Knees et al. ein handelsübliches 5.1-Surround-System [KPSW07b] zur Wiedergabe. Eine noch feinere Spatialisierung ermöglicht die Princeton Display Wall mit 16 um den Benutzer angeordneten Lautsprechern, für die Tzanetakis eine Explorationsumgebung in [TC01] vorstellt.

Beschränkt man sich auf die Simulation einer virtuellen akustischen Umgebung für einen einzelnen Anwender, der einen Kopfhörer trägt, lassen sich Klangquellen aufgrund der getrennten

Signal-Zuführung auf beide Ohren ohne größeren Aufwand mit ähnlicher Ortsauflösung positionieren.

Aus der Hörpsychologie ist bekannt, dass die Ortung eines Klanges durch den Menschen maßgeblich auf drei Faktoren basiert [Kut04]:

- *Interaurale Laufzeitdifferenz*

Der von einer Quelle ausgehende Schall benötigt mitunter unterschiedlich lange, um die beiden Ohren des Hörers zu erreichen. So erreichen die Schallwellen einer Quelle, die sich links vom Hörer befindet, sein linkes Ohr früher als das rechte. Die Auswertung von Phasendifferenzen (primär bei tiefen Frequenzen) und Gruppenlaufzeit-Unterschieden (vor allem bei hohen Frequenzen) ermöglicht die Bestimmung einer seitlichen Schalleinfallsrichtung. Aufgrund der Analyse interauraler Laufzeitdifferenzen allein lässt sich jedoch keine Positionierung der Schallquelle in der Medianebene festmachen: Eine Unterscheidung zwischen Klängen, die den Hörer von vorn, hinten, oben oder unten erreichen, ist so nicht möglich.

- *Interaurale Pegeldifferenz*

Abhängig vom Ursprung der Schallquelle erreichen die Wellen die Ohren des Hörers nicht nur zeitversetzt, sondern sind durch Kopf und Rumpf abgeschattet und gedämpft. So ist der Pegel einer Schallquelle, die sich rechts vom Hörer befindet, auf dem linken Ohr deutlich geringer als auf dem rechten. Diese (frequenzabhängige) Pegelreduktion unterstützt die Bestimmung der seitlichen Einfallsrichtung.

- *Außenohr-Resonanzen*

Das Außenohr des Menschen (bestehend aus Ohrmuschel und dem Beginn des äußeren Gehörganges) führt zu einer Verstärkung (Resonanz) oder Abschwächung bestimmter Frequenzen in Abhängigkeit von der Schalleinfallsrichtung in der Medianebene. Aufgrund der Individualität der Außenohr-Form müssen Details dieses Hörer-spezifischen Zusammenhangs im Laufe der Entwicklung des Menschen gelernt werden.

Ergänzend zu diesen Faktoren, die die Ermittlung der *Richtung* eines Schallereignisses ermöglichen, schätzt der Hörer auf der Grundlage des Pegels und weiterer Klangfarbenveränderungen (wie z.B. einer steigenden Abdämpfung entfernter Quellen, die sich aus dem höheren Energieverlust hochfrequenter Signale in der Luft ergibt) die *Entfernung* der Schallquelle. Die Entfernung-spezifische Pegelreduktion des Tonsignales wird im folgenden Abschnitt „Wiedergabepegel von Schallquellen“ diskutiert.

Die Richtungs- und Ohr-spezifische Signalveränderung einer Schallquelle durch interaurale Laufzeit- und Pegeldifferenzen sowie Außenohr-Resonanzen lässt sich zusammenfassend durch lineare Filter modellieren, die sog. *Außenohr-Übertragungsfunktionen* realisieren. Diese auch als *Head Related Transfer Functions* (HRTF) bezeichneten Funktionen, die sich z.B. durch aufwändige Messungen mit Kunstköpfen ermitteln lassen, ermöglichen für ein beliebiges Tonsignal die Umwandlung in binaurale Nutzsignale, die dem Kopfhörer des Anwenders zugeführt werden können und so die Illusion der Herkunft aus einer vorgegebenen Richtung erzeugen.

Das Klangmodul von Microsofts XNA-Framework dagegen verteilt ein Tonsignal lediglich mit unterschiedlichem Pegel auf die zwei Kanäle einer stereophonen Wiedergabe und ermöglicht somit nur eine sehr eingeschränkte Ortung des Klanges. Aufgrund der beschriebenen

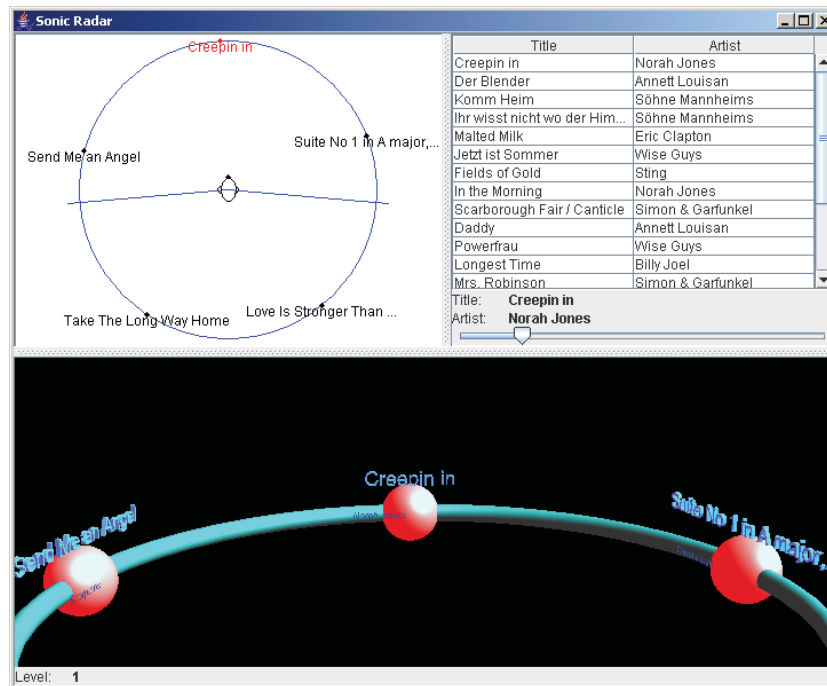


Abbildung 3.21: Bildschirmkopie des SonicRadar-Prototypen

Bedeutung der akustischen Wiedergabe im soniXplorer wurde daher ein OpenAL¹-basiertes Subsystem integriert, das volle HRTF-Unterstützung bietet und somit eine deutlich feinere Lokalisation erlaubt.

Wiedergabepegel von Schallquellen

Die Realisierbarkeit akustischer Orientierung in einer Umgebung gleichzeitig erklingender Musiksingale wurde zunächst in einem Versuchs-System, dem *SonicRadar*, untersucht und in [Lü05] veröffentlicht.

Abbildung 3.21 zeigt eine Bildschirmkopie dieses Prototypen. Grundlage der Visualisierung und Auralisierung ist die hierarchische Partitionierung einer Musikkollektion durch iterative Anwendung von k -Means-Clustering. Für ein vorgegebenes k entsteht so auf jeder Hierarchie-Ebene eine Einteilung in k Gruppen perzeptuell ähnlicher Stücke. Grundlage hierfür ist die Repräsentation von Titeln als Vektoren, die sich aus einer multidimensionalen Skalierung einer Logan/Salomon-basierten Distanzmatrix (vergl. Abschnitt 2.1.1) ergeben. Jedes Cluster wird durch einen Prototyp-Titel repräsentiert, der den geringsten Abstand zum Cluster-Zentrum aufweist.

Abbildung 3.21 stellt die äquidistante Anordnung von fünf Prototypen der obersten Hierarchie-Stufe auf einem Ring dar. Die rechte Tabelle führt alle Titel auf, die sich im aktuell frontal sichtbaren Cluster mit des Prototypen „Creepin in“ befinden. Der Benutzer hat durch einfach Tastenkommandos die Möglichkeit, die Hierarchie-Ebene zu wechseln und so z.B. die Parti-

¹<http://connect.creativelabs.com/openal>

tionierung dieser Stücke in weitere 5 Cluster zu betrachten. Ein Drehen des Ringes ermöglicht ferner, andere Stücke der aktuellen Ebene ins direkte Blickfeld zu bewegen.

In einem ersten Versuch wurde diese visuelle Präsentation der Titel mit der simultanen Wiedergabe aller k Stücke des Ringes gekoppelt, die entsprechend ihrer relativen Position zum Anwender stereophon spatialisiert wurden. Hier zeigte sich jedoch, dass die gleichzeitige Wiedergabe aller Songs zu einem komplexen Klanggemisch führt, das eine Identifikation und Ortung der einzelnen Titel unmöglich macht und somit eher zur Desorientierung des Benutzer beiträgt. Dieses Phänomen ließ sich insbesondere für aktuellere Pop-Songs beobachten, die bereits für sich genommen aus einer Vielzahl unterschiedlicher Klangelemente bestehen.

Eine ähnliche Beobachtung wird von Tzanetakis dokumentiert: Auch die Princeton Display-Wall mit ihren sechzehn dedizierten Lautsprechern ermögliche lediglich die sinnvolle Wiedergabe von acht gleichzeitigen Audio-Strömen [TC01].

Aus diesem Grund erlaubt das SonicRadar die explizite Steuerung eines *Wahrnehmungswinkels* (Angle of Perception), der den Grad der Konzentration auf einen fokussierten Klang (Focus of Perception) bestimmt. Zur Übereinstimmung von akustischer und visueller Wahrnehmung wird stets der frontal sichtbare Teil des Ringes als Fokussierungspunkt angenommen.

Eine Weitung des Wahrnehmungswinkels ermöglicht dem Benutzer somit, sich einen groben Eindruck über die klangliche Umgebung zu verschaffen. Die Einengung des Winkels erlaubt die Konzentration auf frontal positionierte Schallquellen und so in Verbindung mit der Drehmöglichkeit des Ringes eine genauere Untersuchung der k präsentierten Cluster-Prototypen.

Die Ausblendung von Stücken außerhalb des Wahrnehmungswinkels wird durch eine Anpassung des Wiedergabepegels g_{PF} erreicht, der sich wie folgt aus dem Winkel φ zwischen Blickrichtung und dem Vektor vom Standpunkt zur Titel-Position ergibt:

$$g_{PF}(\varphi) = e^{-\frac{\varphi^2}{\sigma^2}}$$

Die Größe des benutzerkonfigurierbaren Wahrnehmungswinkels AoP beeinflusst die Varianz der Exponentialfunktion:

$$\sigma^2 = -\frac{AoP^2}{\ln(gAoP)}$$

Diese Umrechnung stellt sicher, dass Titel am Rand des Wahrnehmungswinkels eine Pegelabschwächung von $gAoP$ erfahren. Im *soniXplorer* ist dieser Parameter auf niedrige $gAoP = 0,1$ gesetzt, um eine effektive Ausblendung der Wiedergabe zu erreichen.

Abbildung 3.22 zeigt den φ -abhängigen Pegelverlauf bei engem ($AoP = 20^\circ$), mittlerem ($AoP = 90^\circ$) und weitem ($AoP = 270^\circ$) Wahrnehmungswinkel. Letztere Konfiguration erlaubt damit das (alltägliche) Hören über den sichtbaren Bereich hinaus.

Aufgrund der positiven Erfahrungen mit dieser Möglichkeit zur akustischen Wahrnehmungsreduktion ist die AoP -abhängige Pegel-Anpassung auch im *soniXplorer* integriert. Um eine Konsistenz zwischen visuellen und akustischen Darstellung sicherzustellen, bedeckt der *soniXplorer* Bereiche außerhalb des Wahrnehmungswinkels mit einem „schwarzen Schleier“. Abbildung 3.23 zeigt diese Abdunklung für die oben genannten Winkelwerte aus einer Vogelperspektive auf die Landschaft.

Ein weiterer Faktor, der den Wiedergabepegel eines Titels beeinflusst, ist die Distanz zur aktuellen Position des Benutzers. Im *soniXplorer* wird ein inverses Distanz-Modell verwendet,

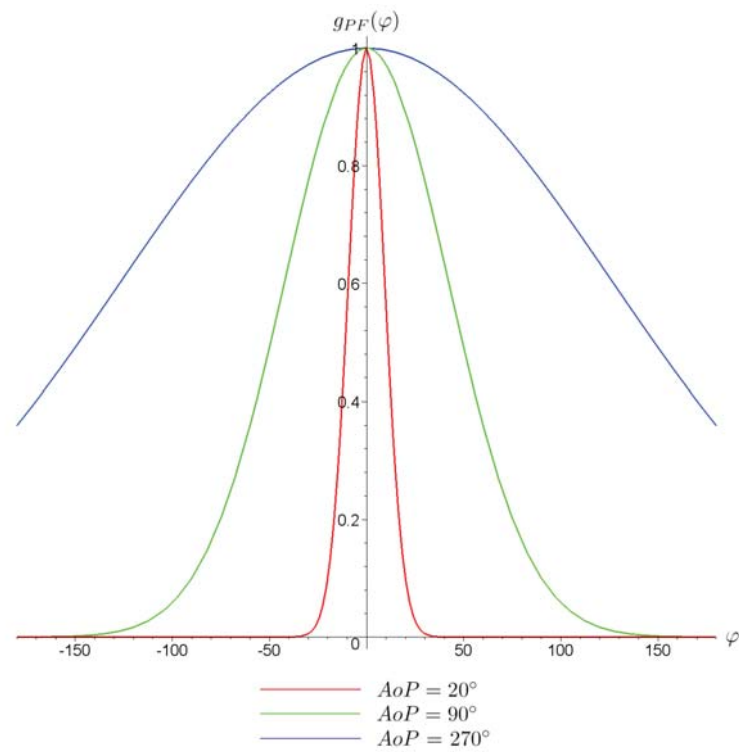


Abbildung 3.22: Wiedergabepegelverlauf $g_{PF}(\varphi)$ bei Wahrnehmungswinkel $AoP = 20^\circ$, $AoP = 90^\circ$ und $AoP = 270^\circ$

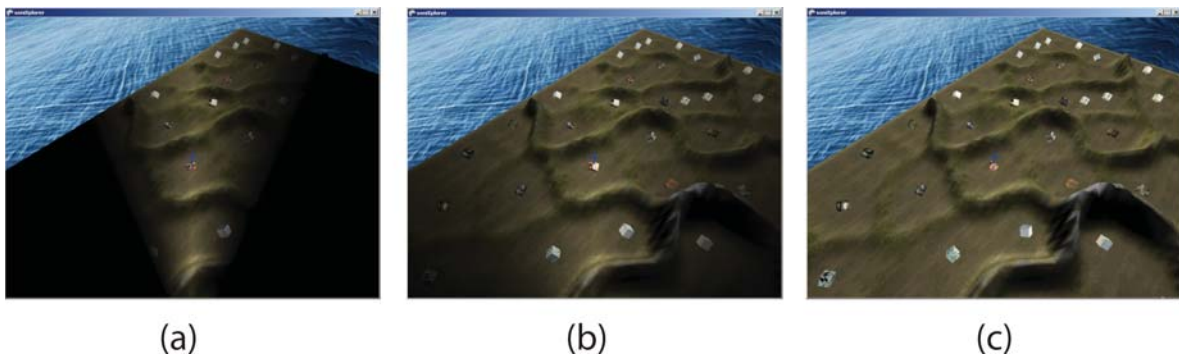


Abbildung 3.23: Visuelle Repräsentation einer Einschränkung des Wahrnehmungswinkels von (a) $AoP = 20^\circ$, (b) $AoP = 90^\circ$ und (c) $AoP = 270^\circ$

das den Pegel g_{Dist} in Abhängigkeit von der Entfernung d wie folgt berechnet:

$$\tilde{g}_{Dist}(d) = \frac{1}{d} \cdot decSpeed - \frac{decSpeed}{minDist} + 1$$

$$g_{Dist}(d) = \begin{cases} 1 & \text{falls } \tilde{g}_{Dist}(d) \geq 1 \\ \tilde{g}_{Dist}(d) & \text{sonst} \\ 0 & \text{falls } \tilde{g}_{Dist}(d) \leq 0 \end{cases}$$

Hierbei bezeichnet $minDist$ die Grenzdistanz, ab der das Signal keine Pegelreduktion mehr erfährt ($\tilde{g}_{Dist}(minDist) = 1$). $decSpeed$ beschreibt die Geschwindigkeit des Pegelabfalls pro Entfernungseinheit und ist ebenfalls benutzerkonfigurierbar. Zur Auralisierung der in den obigen Abschnitten vorgestellten selbstorganisierenden Karte aus 50×80 Zellen wurden Werte von $minDist = 2$ und $decSpeed = 3$ verwendet.

Aufgrund der generierten Hügellandschaft ist es möglich, dass Stücke in der Umgebung der Hörerposition durch Landschaftserhebungen verdeckt sind. In der Realität führt dies zu einer starken Reduktion der Hörbarkeit der Quelle, da nur der um den Berg gebeugte Schall wahrnehmbar ist. Diesem Abschattungseffekt wird im **soniXplorer** durch eine akustische Dämpfung des Pegels durch den Faktor g_{Muffle} Rechnung getragen. Zusätzlich simuliert ein Highcut-Filter die stärkere Pegelreduktion hoher Frequenzen.

Somit berechnet sich die Wiedergabeintensität eines Musikstückes s an Position $\vec{p}_s$ aus der Perspektive eines Hörers an Position $\vec{p}$, der in Richtung $\vec{v}$ schaut, wie folgt:

$$g(s, \vec{p}) = g_{Dist}(\|\vec{p} - \vec{p}_s\|) \cdot g_{PF}(\angle(\vec{v}, \vec{p}_s - \vec{p})) \cdot g_{Muffle}(\vec{p}_s, \vec{p})$$

Während der Evaluation des **soniXplorers** zeigte sich, dass die simultane Wiedergabe von klassischen Stücken und Rockmusik problematisch ist: Aufgrund der deutlich höheren Dynamik klassischer Aufnahmen sind diese im Mittel signifikant leiser und werden von auf Lautheit hin produzierten Rock-Stücken übertönt. Zur Kompensation dieses Effektes wird eine Pegel-Angleichung durchgeführt, die den RMS-Wert (vergl. Abschnitt 2.1.2) wiederzugegebener Titel auf ein einheitliches Niveau anhebt.

Zusammenfassend ergibt sich der Pegel eines Musikstückes s an Position p daher als gewichtete Wiedergabeintensität:

$$\hat{g}(s, \vec{p}) = g_{RMS}(s) \cdot g(s, \vec{p})$$

Auswahl simultan wiedergegebener Schallquellen

Die Pegelreduktion von Stücken, die sich außerhalb des konfigurierbaren Wahrnehmungswinkels befinden, stellt eine erste Reduktion der Komplexität der akustischen Wiedergabe dar. Zur Sonifikation von Kartenregionen, in denen Stücke dicht an dicht positioniert sind und daher mit ähnlichem Pegel wiedergegeben werden würden, reicht dies jedoch nicht aus. Denn die simultane $\hat{g}$ -gewichtete Wiedergabe aller Titel in der Umgebung des Hörers stellt nicht nur hohe Anforderungen an Rechner-Ressourcen, sondern überfordert ferner die Wahrnehmungsfähigkeit des Hörers: Zwei Schallquellen, die aus der gleichen Richtung ertönen, verschmelzen zu einem Höreindruck. Der Benutzer hat damit keine Möglichkeit mehr, die Stücke zu differenzieren und erkennt weder das eine noch das andere.

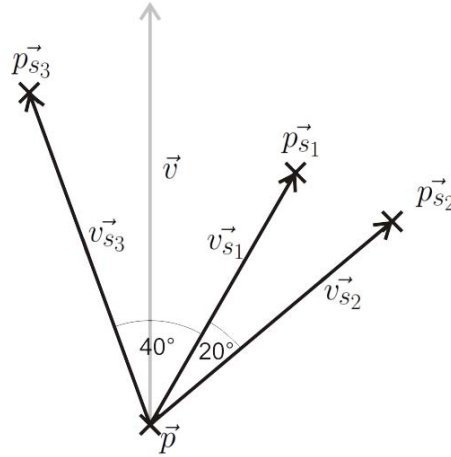


Abbildung 3.24: Skizze zur optimierten Stückauswahl bei drei Titeln

Um dies zu vermeiden, wird im soniXplorer eine Randbedingung für die Auswahl wiedergegebener Stücke eingeführt: Es sei $\vec{v}_s$ der Vektor vom aktuellen Standpunkt $\vec{p}$ des Benutzers zur Position $\vec{p}_s$ des Stückes s .

$$\vec{v}_s := \vec{p}_s - \vec{p}$$

Eine Menge wiedergegebener Titel S muss folgende Anforderung erfüllen:

$$RB : \quad \angle(\vec{v}_s, \vec{v}_{s'}) \geq \minAngle \quad \forall s, s' \in S \text{ und } s \neq s'$$

$\minAngle$ bezeichnet hierbei den Mindestwinkel, der zwischen zwei verschiedenen Stücken der Wiedergabemenge liegen muss. Im soniXplorer wurde empirisch $\minAngle = 60^\circ$ ermittelt; dies bedeutet, dass maximal 6 Stücke gleichzeitig aktiv sind.

Neben dieser obligatorischen Randbedingung lassen sich zwei Qualitätskriterien für eine betrachtete Menge S formulieren:

1. Die Intensität eines Stückes $g(s, \vec{p})$ reflektiert die Bedeutung von s in der aktuellen Hörposition. S sollte die Titel mit dem höchsten Intensitätswert umfassen.
2. Je mehr Stücke (differenzierbar) simultan präsentiert werden, umso informierter und schneller kann sich der Hörer in der Explorationsumgebung bewegen. Daher sollte S möglichst groß sein.

Diese beiden Optimierungsziele lassen sich mitunter nicht in gleichem Maße erfüllen. Ein einfaches Beispiel hierfür ist in Abbildung 3.24 skizziert, die die Positionen von drei Stücken s_1, s_2 und s_3 um den Hörerstandpunkt $\vec{p}$ darstellt. Für die Intensitäten der Titel gelte:

$$g(s_2, \vec{p}) = g(s_3, \vec{p}) = \frac{2}{3}g(s_1, \vec{p})$$

Verfolgt man isoliert die erste Strategie, wählt man zunächst das Stück mit dem größten Intensitätswert s_1 aus. Die Randbedingung RB verhindert dann jedoch die Aufnahme weiterer Titel.

Stellt man das erste Qualitätskriterium zurück, lässt sich mit $\{s_2, s_3\}$ eine Stückselektion finden, die *RB* erfüllt, aber einen Titel mehr umfasst. Bei obigen Intensitätswerten ist die Gesamt-Intensität $g(s_2, \vec{p}) + g(s_3, \vec{p})$ dieser Menge zudem höher als bei der Wahl von s_1 allein. Dies motiviert die folgende Formulierung eines globalen Optimierungskriteriums. Es bezeichne $\mathcal{S}$ die Menge aller Stücke der Kollektion und *thres* einen benutzerdefinierbaren Radius. Um die Berechnungskomplexität zu verringern, schränkt man sich auf die Betrachtung von Stücken in der unmittelbaren Umgebung des Hörers ein:

$$\mathcal{U}(\vec{p}) := \{s \in \mathcal{S} \mid \|\vec{p} - \vec{p}_s\| \leq \text{thres}\}$$

Als global optimal wird die Teilmenge dieser Umgebung betrachtet, deren Intensitätssumme unter Erfüllung der Randbedingung maximal ist:

$$\underset{\substack{S \subseteq \mathcal{U}(\vec{p}) \\ S \text{ erfüllt } RB}}{\operatorname{argmax}} \sum_{s \in S} g^k(s, \vec{p})$$

Durch die Wahl des Parameters k lässt sich ein überproportionaler Einfluss von Stücken mit hohen Intensitätswerten auf die Gesamtsumme realisieren. Höhere Werte für k führen somit zu einer bevorzugten Aufnahme solcher Stücke in die optimale Titelselektion. Damit lässt sich die Relevanz des ersten Qualitätskriteriums für den globalen Optimalitätsbegriff gewichten.

Während der Evaluation des *soniXplorers* wurden unterschiedliche Werte für k erprobt. Es zeigte sich jedoch durchgehend, dass die Auswahl von Stücken nach obigem Kriterium mitunter schwer nachvollzogen werden konnte. Anwender äußerten mehrfach ihre Verwunderung über die Tatsache, dass ein Stück, dem sie (korrekterweise) den höchsten Intensitätswert zuschrieben, nicht akustisch wiedergegeben wurde.

Aufgrund dieser Beobachtung wurde die globale Optimierung zu Gunsten einer lokalen Greedy-Strategie fallen gelassen: Als intuitiver bewerteten die Benutzer ein Verfahren, das die Titel der Umgebung zunächst absteigend nach Intensitätswerten ordnet und Stücke dieser Liste sukzessive in die Wiedergabemenge aufnimmt, so lange sie die Randbedingung nicht verletzen.

Zur expliziten Wiedergabe eines Stückes, das vom *soniXplorer* nicht ausgewählt wurde, können mehrere Strategien verfolgt werden:

- Die Verkleinerung des Wahrnehmungswinkels und die Ausrichtung der Blickrichtung auf den gewünschten Titel führen zu einer Reduktion der Intensitätswerte der anderen Stücke. Daraus resultiert eine automatische Verdrängung dieser Titel aus der Wiedergabemenge zu Gunsten des fokussierten Stückes.
- Der *soniXplorer* erlaubt die explizite benutzergesteuerte Aufnahme von selektierten Stücken in die Wiedergabemenge. Diese Aufnahme erfolgt vor der Betrachtung der sortierten Titelliste und führt somit aufgrund der Durchsetzung der Randbedingung möglicherweise zu einer Verdrängung bereits erklingender Stücke.

Ebenso kann der Anwender ausgewählte Stücke explizit aus der Wiedergabemenge ausschließen.

- Zur isolierten Wiedergabe eines selektierten Titels implementiert der *soniXplorer* einen Solo-Modus (vergl. Abbildung 3.25). Dieser Modus realisiert die aus Standard-Media-Playern bekannte Funktionalität und ermöglicht neben der Konzentration auf ein einzelnes Stück das Vor- und Zurückspulen im Stückverlauf.



Abbildung 3.25: soniXplorer im Solo-Modus zur isolierten Wiedergabe eines markierten Stückes

Um ein unnatürliches plötzliches Erscheinen und Verschwinden von Klangquellen zu vermeiden, blendet der soniXplorer bei Veränderungen der Wiedergabemenge neue Stücke sanft ein und alte Titel langsam aus. Diese graduelle Aktivierung wird durch eine Verringerung der Drehgeschwindigkeit der korrespondierenden Würfel visualisiert.

Außerdem wird bei einer Reaktivierung einer Klangquelle die Wiedergabe an der zuvor unterbrochenen Stelle fortgesetzt. Gerät ein Stück durch die Bewegung des Benutzers aus der Wiedergabemenge und kehrt der Anwender anschließend zur alten Position zurück, erleichtert dies ein akustisches Wiedererkennen der Umgebung.

3.4.3 Navigation

Die Verwendung von Microsofts XNA-Gaming-Engine erlaubt die einfache Abfrage eines angeschlossenen Xbox 360-Controllers. Mit Hilfe dieses Controllers kann sich der Benutzer ähnlich einem Computerspiel in der virtuellen Landschaft bewegen.

Zentrale Steuerelemente für die Navigation sind die in Abbildung 3.26 gekennzeichneten analogen Joysticks, die typischerweise mit dem linken und rechten Daumen bedient werden. Eine völlig ungebundene Beweglichkeit des Benutzers im Raum, die ein Fortschreiten und Drehen in allen Freiheitsgraden erlaubt, erwies sich nicht nur in der Steuerung als zu komplex, sondern führte ferner in initialen Experimenten des öfteren zum Verlust der Orientierung in der Landschaft. Die Unnatürlichkeit solch uneingeschränkter Navigationsmöglichkeiten (insbesondere ohne spürbares Gravitationsfeld) erschwert darüber hinaus die Immersion in die Umgebung.

Daher wurde im soniXplorer eine Bewegungssteuerung implementiert, die der Navigation in klassischen Video-Spielen ähnelt, in denen aus der Sicht eines virtuellen Ichs eine umgebende



Abbildung 3.26: Xbox 360-Controller zur Steuerung des soniXplorers

Landschaft erforscht werden kann.

So wird der Benutzer aufrecht stehend in einem zunächst konstanten geringen Abstand über dem Terrain der aktuellen Position fixiert. Mit Hilfe des rechten Joysticks lässt sich die Blickrichtung nach oben/unten bzw. rechts/links verändern. Eine Vertikalbewegung des linken Joysticks lässt den Betrachter vorwärts bzw. rückwärts gehen. Auf eine Trennung zwischen Blick- und Laufrichtung wurde zur Vereinfachung der Steuerung verzichtet. Daher ist die Horizontalachse des linken Joysticks mit der des rechten Joysticks gekoppelt.

Wie sich während der Entwicklung des ersten Explorations-Prototypen herausstellte, trägt die Verfügbarkeit einer Übersichtsdarstellung der Landschaft maßgeblich zur Orientierung des Benutzers in der Umgebung bei [Nei07]. Neis ergänzt die virtuelle dreidimensionale Umgebung daher um eine zweidimensionale Kartendarstellung, die an die SOM-Visualisierungen von Mörchen oder Pampalk erinnert.

Der soniXplorer ermöglicht eine stufenlose Überblendung zwischen der Terrain-Erkundung in Bodenhaftung und einer Darstellung aus der Vogelperspektive: Mit Hilfe der durch die Zeigefinger bedienbaren Trigger-Tasten des Xbox 360-Controllers lässt sich ein zusätzlicher Abstand zwischen Beobachterposition und Landschaftshöhe einstellen. Dies ermöglicht die Betrachtung der Landschaft in einem Schweben auf einer flexiblen Höhe über dem Terrain.

Abbildung 3.27 illustriert den Blick auf die gleiche Landschaft aus unterschiedlichen Höhen. Während die bodennahe Perspektive eine klarere optische und akustische Trennung zwischen Stück-Clustern realisiert, kann sich der Benutzer durch ein intuitives Aufsteigen einen schnellen Überblick über die dargestellte Kollektion verschaffen.

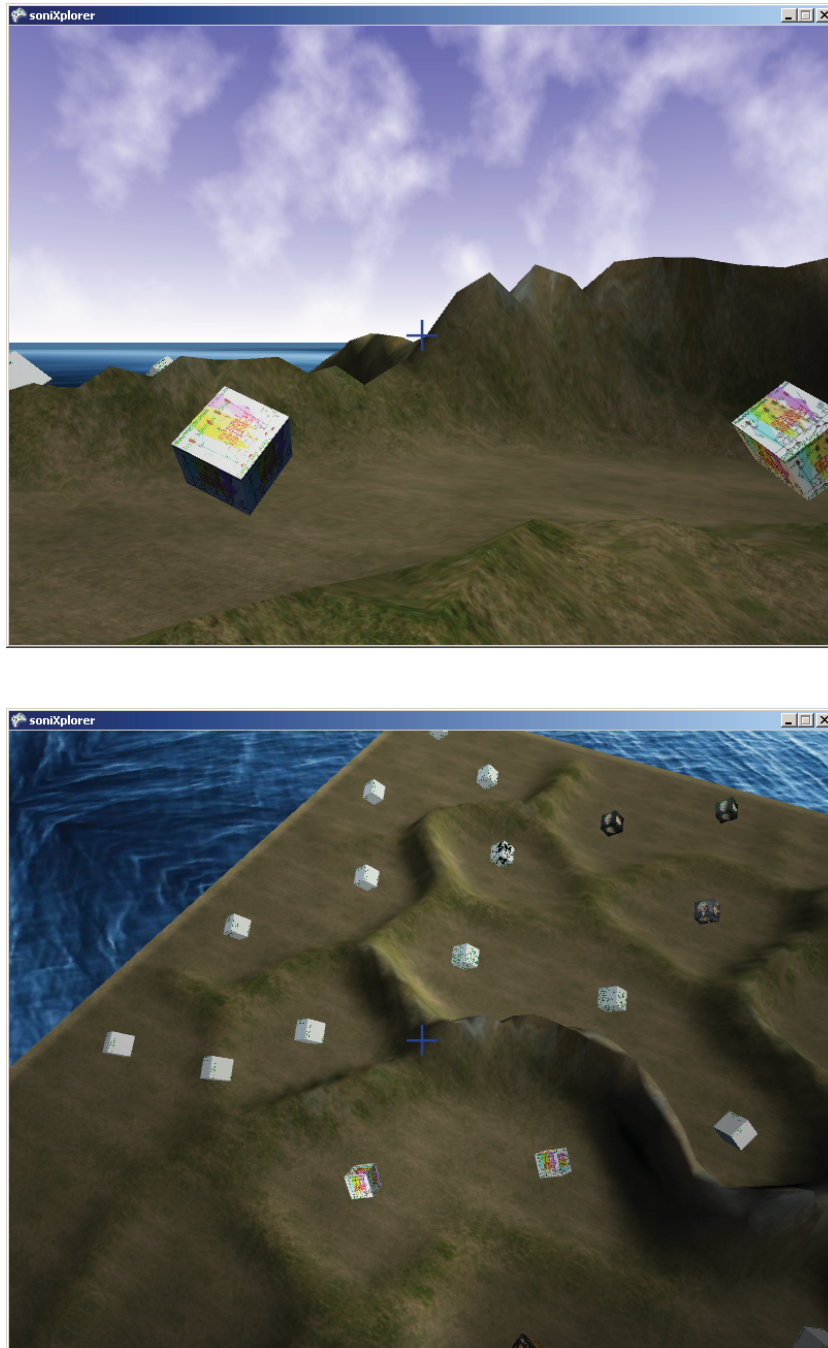


Abbildung 3.27: soniXplorer-Blick auf eine Musikkollektion aus verschiedenen Höhen

Kapitel 4

Interaktion und Nutzer-Adaption

So subjektiv die Wahrnehmung und Beurteilung von Musik ist, so individuell organisieren Benutzer ihre Musiksammlung [CJJ04]. Eine Explorationsumgebung, die einen intuitiven Zugang zu einer Musikkollektion realisieren will, muss diese Nutzer-spezifische Strukturierung berücksichtigen.

Einen ersten Ansatz hierzu leisten die in Abschnitt 2.6 vorgestellten Verfahren, die eine explizite Gewichtung unterschiedlicher musikalischer Aspekte in einem ganzheitlichen Ähnlichkeitsmaß erlauben. Diese Gewichtung muss jedoch vom Benutzer selbst vorgenommen werden und setzt daher eine Kenntnis der Bedeutung der einzelnen Ähnlichkeitsmaße voraus.

Im *soniXplorer* wird daher eine andere Strategie verfolgt: Aus der Beobachtung der Interaktion des Benutzers mit der Explorationsumgebung wird auf das zugrunde liegende Ähnlichkeitsverständnis geschlossen und eine automatische Gewichtung der Teilaspekte vorgenommen.

Diese automatische Adaption des Systems an seinen Anwender ermöglicht die Positionierung neuer und die Repositionierung freigegebener Stücke gemäß dem individuellen Organisationsprinzip des Benutzers. Die so erreichte Simulation eines persönlichen Musikassistenten geht damit deutlich über die Funktionalität bisheriger immersiver Explorationsumgebungen hinaus und gestattet eine bislang nicht erreichbare Vertrautheit des Anwenders mit dem System.

Dieses Kapitel gliedert sich in zwei Hauptteile: Der erste Abschnitt beschreibt Möglichkeiten der Anpassung der Landschaft *durch* den Benutzer. Hierunter fallen das Setzen von optischen Orientierungspunkten, die Verschiebung von Stücken in der Umgebung sowie die Veränderung des Terrain-Höhenprofils.

Im anschließenden Unterkapitel werden Verfahren zur Anpassung der Explorationsumgebung *an* den Benutzer diskutiert. Hier wird beschrieben, wie auf der Grundlage einer Analyse der Anwender-Interaktion eine Neu-Gewichtung von Teilaspekten zum Aufbau eines adaptiven Ähnlichkeitsmaßes vorgenommen wird und wie sich Vektor-Repräsentation und Positionierung eines Stückes in der Landschaft an die neue Distanzfunktion anpassen lassen.

4.1 Interaktion

Die Anpassungsmöglichkeiten der Explorationsumgebung, die dem Benutzer im *soniXplorer* zur Individualisierung der Landschaft zur Verfügung stehen, lassen sich in drei Gruppen unterteilen: Die Veränderung der Stück-Darstellung (insbesondere die Repositionierung eines

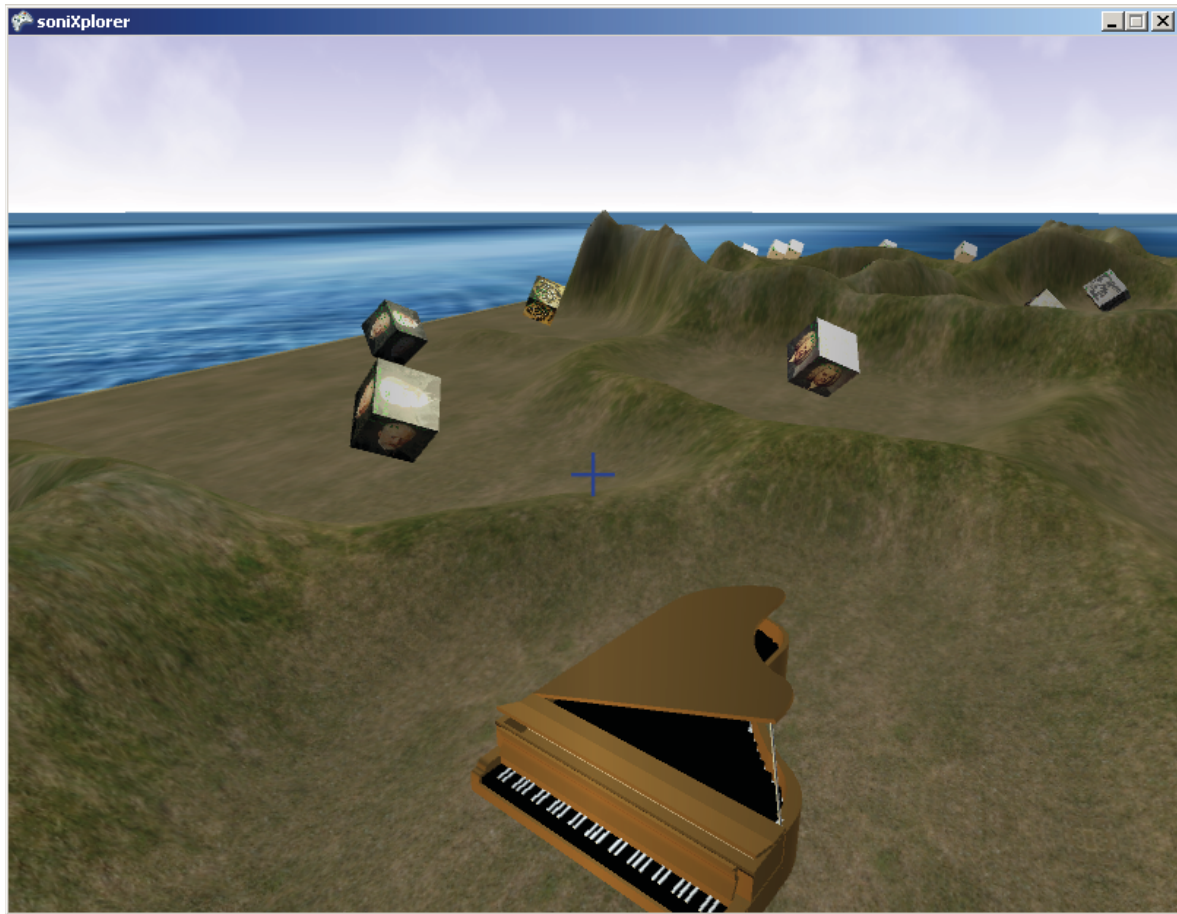


Abbildung 4.1: Visuelle Repräsentation eines Stückes als Flügel-Symbol

Titels in der Landschaft), das Setzen von konfigurierbaren Hinweis-Schildern, die als Orientierungspunkte in der Landschaft fungieren, und die Modifikation des Terrain-Höhenprofils, d.h. die Erhöhung oder Absenkung von trennenden Bergketten.

4.1.1 Anpassung der Stück-Darstellung

Die optische Repräsentation von Stücken in der Landschaft ist standardmäßig ein schräggestellter Würfel, der mit dem Cover-Bild des zugehörigen Albums texturiert ist. Zur suggestiven grafischen Kennzeichnung lässt sich diese Form durch ein beliebiges 3D-Modell ersetzen, das den Titel treffender visualisiert.

Abbildung 4.1 zeigt beispielhaft einen Blick auf die Stücke 413, 414, 303 und 324 der Klassik-Region der Beispielsammlung aus Abschnitt 3.4 (vergl. insbesondere Abbildung 3.13), in der der 2. Satz aus dem 9. Klavierkonzert von Wolfgang Amadeus Mozart (302) im Vordergrund durch ein Flügel-Symbol dargestellt ist.

Entspricht die mit Hilfe der selbstorganisierenden Merkmalskarte gefundene Gruppierung von Stücken nicht dem Strukturierungsprinzip des Benutzers, so hat er die Möglichkeit, einzelne Titel auf andere Positionen der Landschaft zu verschieben. Hierzu markiert er zunächst das

betreffende Stück, bewegt sich anschließend in die Zielregion und legt den Titel schließlich an der mit einem mittigen Fadenkreuz angepeilten Position ab.

Dieses Verfahren erwies sich gegenüber einem aus der Mausbedienung vertrauten Drag'n'drop-Ansatz aus folgenden Gründen als vorteilhaft:

- Ein Drag'n'Drop-Ansatz erfordert ein Halten einer Controller-Taste während des gesamten Verschiebe-Vorgangs und erschwert somit ein Repositionierung in entfernte Regionen.
- Die der Repositionierung vorausgehende freie Navigation in der Landschaft ermöglicht eine konzentrierte Auswahl einer Ziel-Region ähnlicher Titel.

Wie sich bei der Beobachtung der Evaluationsteilnehmer zeigte, ist dieses Verfahren insbesondere gut mit der Möglichkeit des Abhebens vom Boden zum Wechsel in eine Art Vogelperspektive auf die Landschaft kombinierbar.

Neben der expliziten Verschiebung eines Stückes an eine neue Position bietet der **soniXplorer** dem Anwender die Möglichkeit, einen Titel von seiner jetzigen Platzierung zu lösen und ihn im nächsten Anpassungsschritt (vergl. Abschnitt 4.2) auf der Grundlage eines adaptierten Ähnlichkeitsbegriffes in einer anderen Region automatisch positionieren zu lassen. Dies ermöglicht dem Benutzer, seine Unzufriedenheit mit der aktuellen Platzierung zu kommunizieren, ohne eine konstruktive Alternativposition angeben zu müssen.

Die „Freisetzung“ eines Titels wird optisch durch eine verblasste halb-transparente Darstellung des Stück-Modells angedeutet (siehe Abbildung 3.19).

4.1.2 Orientierungspunkte

In initialen Experimenten zeigte sich, dass trotz der Möglichkeit zur überblicksartigen Darstellung aus der Vogelperspektive die Benutzer nach mehrmaligen Drehbewegungen aufgrund der visuellen Gleichartigkeit des Terrains hin und wieder die Orientierung in der Landschaft verloren.

Zur klaren Kennzeichnung von Regionen ermöglicht der **soniXplorer** daher die Platzierung von weit sichtbaren Schildern, die als Orientierungspunkte zur Beschreibung eines Terrain-Bereiches genutzt werden können.

Prototypisch wurden drei verschiedene Formen implementiert. Abbildung 4.2 zeigt diese 3D-Modelle in der Landschaft. Zur individuellen Markierung lassen sich die Flächen mit beliebigen Texturen belegen und mit frei konfigurierbaren Schriftzügen überziehen. Ferner kann die Größe und insbesondere die Stabhöhe der Schilder angepasst werden. Hierdurch wird eine Sichtbarkeit der Orientierungspunkte sogar aus Terrain-naher Position in bergiger Landschaft erreicht.

Mit Hilfe solcher Hinweis-Schilder kann der Benutzer Regionen der Karte individuell annotieren. In einer auf Genre basierenden Stückgruppierung wie in Abbildung 3.13 könnten z.B. diese Genre-Bezeichnungen zur Kennzeichnung der Cluster verwendet werden.



Abbildung 4.2: Texturierte Schilder verschiedener Formen als Orientierungspunkte in der Landschaft

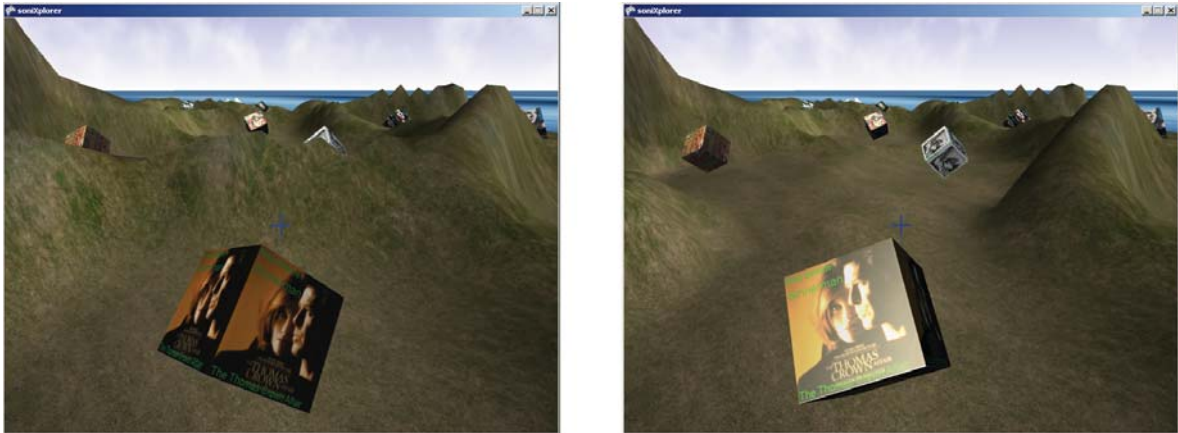


Abbildung 4.3: Explorationsumgebung vor und nach Entfernung einer trennenden Bergkette

4.1.3 Anpassung des Terrains

Die Strukturierung einer Musikkollektion stellt sich für den Nutzer zum einen als gruppierenden Anordnung von Stücken in der Nachbarschaft zu ähnlichen Titeln und zum anderen als Trennung zwischen solchen Clustern homogener Stücke durch Bergketten dar. Zur Verstärkung oder Abschwächung dieser Separierung erlaubt der *soniXplorer* die Veränderung des Höhenprofils der Landschaft.

So könnte der Benutzer die in Abbildung 3.13 erkennbare Abtrennung des Songwriter-Stückes 363 von den anderen Titeln des gleichen Genres durch eine Abflachung der umgebenden Hügel aufheben. Abbildung 4.3 illustriert dies durch eine Darstellung der Umgebung vor und nach einer Anpassung des Höhenverlaufs.

Zur Durchführung von Terrain-Veränderungen stehen dem Anwender zwei Möglichkeiten zur Verfügung:

1. Lokale Absenkungen bzw. Anhebungen der Landschaft lassen sich mit der über den Trigger-Tasten positionierten linken bzw. rechten Shoulder-Taste des Xbox360-Controllers vollziehen, die direkt erreichbar mit dem Zeigefinger bedient werden kann. Hierzu wird die Höhenveränderung auf einen konfigurierbaren kreisförmigen Bereich um den Punkt, der durch das Fadenkreuz in der Mitte der Bildfläche fokussiert wird, ausgedehnt. Die Intensität der Anhebung bzw. Absenkung der Umgebung um Kartenpunkt $\vec{p}_0$ pro Zeitintervall, in dem eine Controller-Taste gedrückt gehalten wird, wird mit Hilfe einer Gauss-Funktion beschrieben:

$$h_{new}(\vec{p}) = h_{old}(\vec{p}) + e^{\frac{||\vec{p}-\vec{p}_0||}{\sigma^2 \cdot resol^2}} \cdot k$$

Hierbei parametrisiert σ die Radius-Größe der beeinflussten Umgebung. Große Werte von σ führen zu einer weiträumigen Veränderung, kleine Werte von σ beschränken die Anpassung auf Kartenpunkte $\vec{p}$ in unmittelbarer Nähe zu $\vec{p}_0$. *resol* bezeichnet den in Abschnitt 3.4.1 beschriebenen Faktor zur Auflösungsverfeinerung, k quantifiziert die (positive oder negative) Veränderung der Berghöhe im Punkt $\vec{p}_0$ pro Zeiteinheit.

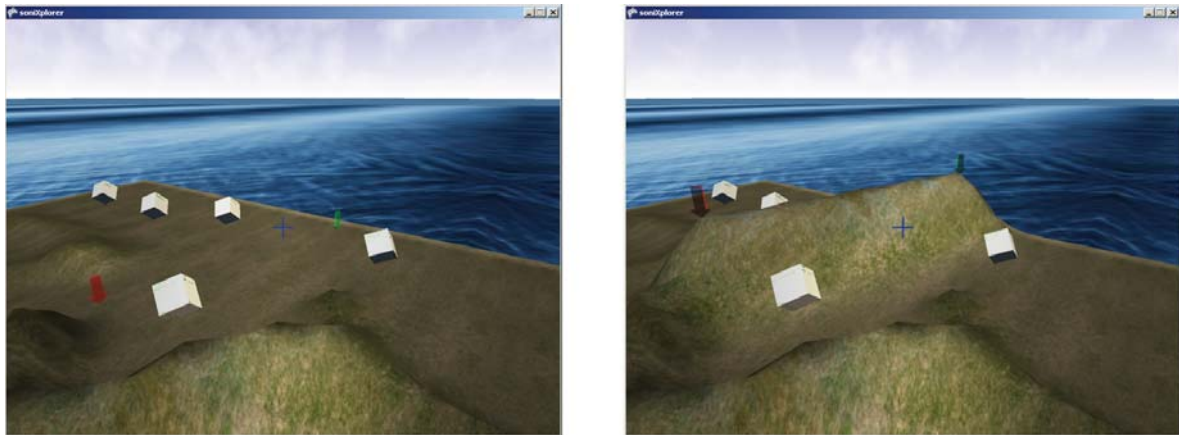


Abbildung 4.4: Aufbau einer zusammenhängenden trennenden Bergkette

2. Zur Veränderung der Höhe zusammenhängender Bergketten erlaubt der *soniXplorer* die Festlegung eines Anfangs- und Endpunktes in der Landschaft. Danach lässt sich mit den gleichen Bedienelementen wie bei der lokalen Anpassung das Niveau der zwischen diesen Marken liegenden Kartenpunkte erhöhen oder erniedrigen. Dies erlaubt den schnellen Auf- bzw. Abbau einer trennenden Bergkette zwischen zwei Kartenregionen. Abbildung 4.4 illustriert dieses Verfahren.

Anpassungen der visuellen Darstellung spiegeln sich naturgemäß auch in der akustischen Präsentation der Landschaft wider, z.B. in Form einer Abschattung von Schallquellen durch zusätzliche Bergketten. Die Aufhebung bzw. Errichtung von Separierungen zwischen Kartenregionen fließt ferner - wie im nächsten Abschnitt beschrieben - in die Anpassung des zugrunde liegenden Ähnlichkeitsmaßes ein.

4.2 Adaption

Die Individualität von Musikgeschmack und die Schwierigkeit, diesen in einer ausführbaren Suchanfrage zu artikulieren, erklären die Popularität der Integration von Benutzer-Feedback-Mechanismen in Empfehlungssysteme für Musik. Während diese zumeist das Ziel einer individuellen Zusammenstellung von Stücken verfolgen, die beim Nutzer Gefallen finden, ist für die personalisierte strukturierte Darstellung einer Musikkollektion eine weitergehende, differenziertere Adaption an den Benutzer erforderlich:

Die Gruppierung von ähnlichen Stücken und die relative Positionierung dieser Cluster zueinander wie sie durch eine selbstorganisierende Merkmalskarte erzeugt werden können, setzen eine Korrelation des verwendeten Distanzmaßes mit dem Ähnlichkeitsempfinden des Benutzers voraus. In diesem Unterkapitel wird daher der im *soniXplorer* eingesetzte Mechanismus zur Personalisierung des Ähnlichkeitsbegriffes durch adaptive Gewichtung der implementierten Teilaspekte von musikalischer Ähnlichkeit diskutiert und seine Auswirkungen auf die Anpassung einer selbstorganisierenden Karte vorgestellt.

Das „Erlernen“ eines Ähnlichkeitsverständnisses aus der Beobachtung der Interaktion des Benutzers mit der Explorationsumgebung verfolgt zwei Zielsetzungen: Zum einen erlaubt es die

nachvollziehbare (und „erwartete“) Positionierung von hinzugefügten Stücken in der Landschaft; zum anderen kann der Benutzer bewusst Stücke, deren Lage er für unintuitiv hält, freisetzen und so eine Neu-Positionierung durch das System auf der Grundlage eines veränderten Ähnlichkeitsbegriffes veranlassen.

4.2.1 Verwandte Arbeiten

Die Literatur zur Personalisierung von Musik-Informationssystemen ist umfangreich und vielfältig. Dieser Abschnitt beschränkt sich daher auf die Diskussion eines heterogenen Querschnitts von relevanten Ansätzen, die durch unterschiedliche Methoden eine Anpassung an den aktuellen Benutzerkontext realisieren.

Anpassung aufgrund physiologischer Parameter

Die aktivierende Wirkung von rhythmischer Musik hat die Begleitung von sportlichen Aktivitäten durch die Wiedergabe Leistungs-motivierender Stücke populär gemacht. Aufgrund der fortschreitenden Miniaturisierung von Abspielgeräten hat diese Begleitung insbesondere in Outdoor-Aktivitäten wie z.B. dem Joggen Einzug erhalten.

Aufgrund der Tendenz des Joggers zur Synchronität mit den Intensitätsschwerpunkten der gehörten Musik fungiert der durchgängige Beat eines wiedergegebenen Stückes dabei als eine Art Taktangeber für das Lauftempo. Damit trägt er nicht nur zur Stabilisierung der Schrittgeschwindigkeit bei, sondern lässt sich bewusst als Impulsgeber zur Beschleunigung oder Verlangsamung des Lauftempos einsetzen.

Oliver und Kreger-Stickles entwickeln auf der Grundlage dieser Beobachtungen ein System zur situativen Musik-Wiedergabe [OKS06]. Basis der Adaption an den aktuellen Kontext ist keine explizite Interaktion des Benutzers mit dem System, sondern die Messung physiologischer Parameter des Läufers. Beispielhaft integrieren sie in ihren Prototyp *PAPA* die Erfassung der Herzfrequenz. Ergänzt wird diese Messung durch die Ermittlung der aktuellen Schrittgeschwindigkeit.

Oliver und Kreger-Stickles konzipieren ihr System als „persönlichen Fitness-Trainer“, der in Abhängigkeit vom aktuellen Anstrengungs-Niveau des Benutzers und einem vorgegeben Workout-Profil Stücke aussucht, die zur Beschleunigung oder Reduktion des Lauftempos (und damit der Herzfrequenz) führen sollen. Hierzu analysiert *PAPA* zunächst mittels automatischer Tempo-Erkennungs-Verfahren (vergl. Abschnitt 2.1.2) die vorgegebene Musikkollektion. Der implementierte Auswahlalgorithmus ist recht geradlinig: Ist die gemessene Herzfrequenz unterhalb dem im Profil vorgegebenen Soll-Wert, wird ein Stück aus der Kollektion ermittelt, dessen Tempo oberhalb der aktuellen Schrittgeschwindigkeit liegt. Die Tempodifferenz korrespondiert dabei mit der Distanz zwischen Ist und Soll der Herzfrequenz. Analog wird bei einer zu hohen Herzfrequenz ein proportional langsames Stück ausgewählt.

Damit folgt *PAPA* nicht nur einem vorab festgelegten Workout-Profil, sondern passt die Wiedergabe von Titeln dem aktuellen Benutzerkontext (gekennzeichnet durch Herz- und Schrittgeschwindigkeit) an.

Die Ideen dieses Prototypen sind inzwischen in einem kommerziellen Produkt der Firma Yamaha realisiert, dem *BodiBeat*-System¹.

¹www.yamaha.com/bodibeat

Dialogisches Klassifikator-Training

Mandel et al. stellen ein inhaltsbasiertes Empfehlungssystem vor, das ein Benutzermodell aus der Reaktion des Anwenders auf bewusst ausgewählte Stücke generiert [MPE06].

Hierzu werden zunächst statistische Kenngrößen der MFCC-Verteilung (vergl. Abschnitt 2.1.1) in den Titeln der betrachteten Kollektion ermittelt und eine Matrix paarweiser Distanzen gebildet. Zur Identifizierung von Stücken, die den Geschmack des Anwenders ansprechen, wird eine Support-Vector-Machine gelernt, die durch eine Hyper-Ebene im Merkmalsraum die empfohlenen von den nicht empfohlenen Titeln trennt. Da die Entscheidungsfunktion eines Support-Vector-Klassifikators ausschließlich auf Skalarprodukten basiert, lässt sich die explizite Parametrisierung dieser Hyper-Ebene durch eine Auswahl von repräsentativen „Stütz-Vektoren“ (support vectors) ersetzen (Dualitäts-Prinzip).

Der Aufbau eines personalisierten Benutzerprofils kann damit auf die geschickte Auswahl von Stütz-Vektoren und deren Beurteilung durch den Anwender zurückgeführt werden. Hierzu setzen Mandel et al. auf Verfahren des „active learning“, die heuristisch die für die Klassifikator-Induktion informativsten Stücke identifizieren. Wie in [MPE06] nachgewiesen, ermöglicht dieser dialogische Adaption eines Empfehlungssystems eine signifikant schnellere Qualitätsverbesserung der generierten Empfehlungen als herkömmliche Relevanz-Feedback-Verfahren, die üblicherweise eine Reaktion des Benutzers auf die zuletzt erzeugte Empfehlungsliste verlangen.

Adaptive Suchanfragen-Reformulierung

Auch das adaptive Musikinformationssystem von Hoashi et al. basiert auf einer Analyse der MFCC-Koeffizienten der betrachteten Titel [HMI03]. Zur Aggregation dieses Fensterbasierten Merkmals wird ein Vektor-Quantisierer eingesetzt [Foo97], der mit Hilfe eines erlernten Entscheidungsbaumes einem Stück einen r -dimensionalen Vektor zuordnet, dessen Komponenten als Histogramm einer Zuordnung zu einer der r Klassen interpretiert werden können. Im Kontext eines Empfehlungssystems fungiert eine Quantisierung des Relevanzwertebereiches als Klassenmenge.

Die Darstellung eines Titels als r -dimensionaler Vektor ermöglicht die Berechnung der Ähnlichkeit zweier Stücke mit Hilfe des Cosinus-Maßes. Bildet man den Mittelwert einer Menge von bevorzugten Stücken, so lässt sich dieses Maß ebenso zur Generierung von Empfehlungen verwenden.

Diese Empfehlungen kann der Benutzer als relevant oder irrelevant kennzeichnen. Auf der Grundlage seiner Klassifizierung lässt sich mit Hilfe des Rocchio-Algorithmus ([Roc71], vergl. Abschnitt 2.4.4) eine Verschiebung des Anfragevektors von den irrelevanten zu den relevanten Stücken realisieren.

Die Anpassung des Systems an den Benutzer besteht bei Hoashi et al. somit in einer Reformulierung der Suchanfrage, die sich aus den Relevanzbewertungen des Anwenders ergibt.

Wie Hoashi et al. in einer späteren Arbeit darstellen, ist die Modellierung eines Benutzerprofils durch die Mittelung der Vektor-Repräsentation bevorzugter Stücke problematisch, insbesondere, wenn der Anwender heterogene Musik-Interessen besitzt [HIMS07]. Sie plädieren in diesem Fall für eine Gruppierung ähnlicher Stück-Präferenzen durch einen Cluster-Algorithmus und die Integration unabhängiger Suchanfragen mit den Zentren dieser Cluster.

Auch mit dieser Erweiterung zeigt sich jedoch ein Nachteil einfacher Empfehlungssysteme gegenüber strukturierten Kollektions-Darstellungen wie dem *soniXplorer*: Generierte Empfehlungen basieren auf einem (möglicherweise heterogenen) Geschmacks-Modell eines Anwenders, lassen aber den weiteren Kontext (wie z.B. die aktuelle Stimmung des Benutzers oder den intendierten Verwendungszweck des Stückes) außer Betracht. Strukturierte Darstellungen dagegen erlauben die bewusste Konzentration des Benutzers auf bestimmte Bereiche seiner Kollektion, die in der aktuellen Situation für ihn relevant sind und ermöglichen somit einen gezielteren (und damit schnelleren) Zugriff auf präferierte Stücke.

Personalisierte Distanz-Kombination

Voraussetzung für eine hilfreiche und intuitiv verständliche Strukturierung ist die Wahl eines geeigneten Ähnlichkeitsmaßes. Wie beschrieben lässt sich Ähnlichkeit nur in Grenzen objektiv quantifizieren; daher ist ein adaptiver Ähnlichkeitsbegriff erforderlich.

In Abschnitt 2.6 wurden bereits Ansätze zur bewusst von Benutzer gesteuerten Kombination von Ähnlichkeitsmaßen diskutiert. Im folgenden werden Verfahren zur automatischen Induktion dieser Kombination aus explizitem oder implizitem Anwender-Feedback vorgestellt.

Semi-Überwachtes Clustering Xing et al. betrachten die Adaption eines Distanzmaßes im Kontext einer Clustering-Aufgabe: Ziel ist die Gruppierung von Stücken unter definierbaren Randbedingungen [XNJR03]. Mit Hilfe dieser kann der Benutzer die Zusammengehörigkeit von Titeln vorgeben und somit spezifizieren, dass der Abstand zwischen diesen Stücken unter der herzuleitenden Distanzfunktion gering sei. Dies ermöglicht einen gezielten und intuitiven Eingriff in die automatische Strukturierung.

Ausgangspunkt sind die Repräsentation eines Titels s_i als n -dimensionaler Vektor $\vec{x}_i$ sowie die Euklidische Distanz $\|\vec{x}_i - \vec{x}_j\|$ zur Quantifizierung der Distanz zwischen zwei Stücken.

Zur Anpassung an die vorgegebenen Randbedingungen verallgemeinern Xing et al. dieses Abstandsmaß zu einer quadratischen Form, die durch die positiv definite Matrix $n \times n$ -Matrix A parametrisiert werden kann:

$$d(\vec{x}_i, \vec{x}_j) = \|\vec{x}_i - \vec{x}_j\|_A = \sqrt{(\vec{x}_i - \vec{x}_j)^T A (\vec{x}_i - \vec{x}_j)}$$

Da $d(\vec{x}_i, \vec{x}_j) = \|A^{1/2}\vec{x}_i - A^{1/2}\vec{x}_j\|$ gilt, kommt die Bestimmung einer Matrix A der Suche nach einer geeigneten linearen Transformation der Stück-Repräsentationen gleich.

Auf dieser Grundlage lässt sich folgendes Optimierungskriterium formulieren:

$$\min_A \sum_{(\vec{x}_i, \vec{x}_j) \in \mathcal{S}} \|\vec{x}_i - \vec{x}_j\|_A^2$$

Gesucht wird somit eine Matrix A , deren zugehöriges Distanzmaß die vom Benutzer vorgegebenen Stück-Paare in $\mathcal{S}$ möglichst dicht zusammenrückt.

Unter Hinzunahme weiterer Randbedingungen lässt sich dieses konvexe Optimierungsproblem mit Hilfe klassischer iterativer Lösungsalgorithmen wie Newton-Verfahren oder Gradient Descent lösen. Xing et al. differenzieren hierzu zwischen dem Fall einer beliebigen positiv definiten Matrix A und einer Diagonal-Matrix, die eine intuitive Interpretation als Gewichtung einzelner Vektor-Dimensionen erlaubt.

Einer Anwendung dieser Distanzmaß-Adaption im Kontext der *soniXplorer* -Explorationsumgebung stehen folgende Gründe entgegen:

- Xing et al. reduzieren die Einflussnahme des Benutzers auf die Festlegung von Paaren zueinander ähnlicher Stücke. Die Repositionierung von Stücken im *soniXplorer* ermöglicht jedoch eine quantitative Bewertung des Abstandes zwischen Stücken, insbesondere auch die bewusste örtliche Trennung von Titeln.
- Das Modell erlaubt keine Berücksichtigung von Terrain-Höhen-Veränderungen.
- Xing et al. kombinieren und gewichten die Koordinaten einer Vektor-Repräsentation der Stücke. Vor der Multidimensionalen Skalierung liegen im *soniXplorer* jedoch nur Matrizen paarweiser Distanzen vor. Die durch die Multidimensionalen Skalierung zugeordneten Koordinaten simulieren zwar die Kombination der Einzelähnlichkeiten, eine Zuordnung der Dimensionen zu musikalischen Teilaspekten geht jedoch verloren. Dies erschwert nicht nur eine Interpretierbarkeit der Matrix A ; die verlorene Restriktion auf einen „sinnhaften“ Suchraum erhöht darüber hinaus die Gefahr einer Überanpassung von A .

Adaptive Organisation von Bild-Gruppen Ein weiteres Beispiel für ein System, das Personalisierung durch eine Anpassung der verwendeten Distanzfunktion realisiert, ist das Multimedia Management-Werkzeug *EGO* von Urban und Jose, das vor allem bei der Organisation von Bildersammlungen Unterstützung bietet [UJ06].

Zur individuellen Strukturierung einer Kollektion kann der Benutzer auf einer großen Arbeitsfläche Bilder zu Gruppen zusammenfassen. Das System unterstützt den Anwender bei der Aufnahme weiterer Bilder in die selektierte Gruppe, indem es dazu passende Objekte aus der Gesamt-Sammlung vorschlägt. Diese Funktionalität wird durch ein Empfehlungssystem implementiert, dessen Ähnlichkeitsbegriff durch die Auswahl tatsächlich hinzuzufügender Bilder angepasst wird.

EGO extrahiert für ein Bild b_i zunächst l inhaltsbasierte Merkmale. Jedes dieser Merkmale wird durch einen mehrdimensionalen Vektor dargestellt. Damit ist b_i durch $\vec{x}_i = (x_{i1}, \dots, x_{il})^T$ beschrieben.

Die Generierung einer Empfehlung auf der Basis von N selektierten Bildern, die als positive Beispiele für das vom Benutzer intendierte Konzept der Bildergruppe betrachtet werden, basiert auf einer zweistufigen Ähnlichkeitsberechnung zu einem Anfragevektor $\vec{q}$, der sich aus der Mittelung der ausgewählten Gruppen-Bilder-Repräsentationen ergibt.

1. Für jedes der l Merkmale wird zunächst eine Einzel-Distanz berechnet:

$$d_k(\vec{q}, \vec{x}_i) = \|\vec{q}_k - \vec{x}_{ik}\|_{W_k} \quad \text{für } k = 1, \dots, l$$

W_k beschreibt hierbei (bis auf eine unwesentliche Skalierung) die Inverse der Kovarianzmatrix der k .ten Merkmalsausprägungen von $\vec{x}_i$ und $\vec{q}$. Damit stellt d_k eine Variante der Mahalanobis-Distanz dar.

2. Die so ermittelten Abstände der l Bild-Merkmale werden linear kombiniert:

$$d(\vec{q}, \vec{x}_i) = \vec{u}^T \cdot (d_1(\vec{q}, \vec{x}_i), \dots, d_l(\vec{q}, \vec{x}_i))^T$$

Die gemäß dieser Funktion ähnlichsten 9 Bilder werden dem Benutzer zur Aufnahme in die selektierte Gruppe vorgeschlagen. Aufgrund der tatsächlichen getroffenen Auswahl wird im Anschluss daran die Ähnlichkeitsfunktion angepasst: Die neu gebildete Bildergruppe führt nicht nur zu aktualisierten *Intra-Merkmal*-Gewichten W_k , sondern auch zu einer Neu-Berechnung des *Inter-Merkmal*-Gewichtsvektors $\vec{u}$. Hierzu wird zunächst die kumulierte Distanz der N' Gruppenbilder vom Abfragevektor bzgl. des Merkmals j ermittelt:

$$cd_j = \sum_{n=1}^{N'} d_j(\vec{q}, \vec{x}_n)$$

Das Gewicht des k .ten Merkmals in der Gesamtdistanz ergibt sich aus dem summierten Vergleich von cd_k mit den kumulierten Distanzen der anderen Merkmale:

$$u_k = \sum_{j=1}^l \sqrt{\frac{cd_j}{cd_k}}$$

Eine statistisch motivierte Herleitung dieser Formeln findet sich in [RH00]. Man erkennt, dass Merkmalen, die einen vergleichsweise kleinen Abstand zum Anfragevektor haben, eine höhere Gewichtung bei der zukünftigen Empfehlungsgenerierung zugewiesen wird.

Distanzmaß-Kombination mit RBFN Als letztes Beispiel für die Verwendung einer adaptiven Distanzfunktion wird im folgenden das Musik-Empfehlungssystem *MUSIPER* von Sotiropoulos et al. vorgestellt [SLT08]. Die Autoren unterstützen die Ansicht, dass unterschiedliche Ähnlichkeitsaspekte für verschiedene Anwender unterschiedlich wichtig, wenn überhaupt verständlich sind. Primäres Ziel der Personalisierung in *MUSIPER* ist daher die Auswahl einer Merkmalsteilmenge, die das Ähnlichkeitsverständnis des Benutzers am besten erklärt. Darüber hinaus wird ein Modell zur Kombination der ausgewählten Features angepasst.

Jedes Stück wird durch N inhaltsbasierte Merkmale charakterisiert. Der Adaptionprozess beginnt mit der Formulierung eines N -dimensionalen Anfragevektors $\vec{q}$, der sich z.B. in einer Query-By-Example-Anfrage aus der Repräsentation eines Titels ergibt. *MUSIPER* betrachtet M verschiedene Teilmengen $C_1, \dots, C_M$ der verfügbaren Merkmale und ermittelt für jedes Stück s_i und jede Teilmenge C_j einen K_j -dimensionalen Vektor, der die komponentenweise Distanz zwischen den Merkmalen von s_i und $\vec{q}$ darstellt. K_j bezeichnet dabei die Anzahl der Merkmale in C_j ($K_j \leq N$).

Grundlage des zusammenfassenden Ähnlichkeitsbegriffes ist damit - im Gegensatz zu den bisher diskutierten Ansätzen - eine Auflistung von Einzel-Distanzen, die sich nicht notwendigerweise aus der Berechnung des Euklidischen Abstandes einer Vektor-Repräsentation ergeben müssen.

Zur Kombination der Einzel-Maße wird jeder Merkmalsteilmenge C_j ein separates Neuronales Netz zugeordnet. Die Topologie dieses Netzes besteht aus K_j Eingabeknoten, die vollständig mit 5 verborgenen Neuronen verknüpft sind. Diese Knoten sind schließlich mit einem Ausgabe-Neuron verbunden. Als Aktivierungsfunktion der inneren Knoten wird eine radiale Basis-Funktion (typischerweise eine Gauss-Funktion konfigurierbarer Breite um ein K_j -dimensionales Zentrum) verwendet. Die Ausgabe dieser verborgenen Neuronen wird linear zur Ausgabe des Netzes kombiniert. Diese *Radial Basis Function Networks* (RBFN) lassen

sich als eine Art instanzbasierte Regressoren begreifen, die zwischen den durch die Zentren der inneren Knoten dargestellten Stützpunkten interpolieren.

Im Verlauf des Adaptionprozesses identifiziert zunächst jedes der M Netze das zu $\vec{q}$ ähnlichste Stück der Kollektion. Der Nutzer bewertet anschließend diese Relevanz (und damit die Ähnlichkeit zu $\vec{q}$) dieser maximal M identifizierten Titel. Dieses Anwenderfeedback wird daraufhin in einem Backpropagation-Verfahren zur Anpassung der Netz-Parameter genutzt. Nach einer vorgegebenen Anzahl von Iterationen entscheidet sich MUSIPER für die im Gesamtverlauf treffendste Merkmalsteilmenge und ordnet gemäß des angepassten RBFNs die Titel der Kollektion absteigend nach Ähnlichkeit an.

Auch wenn Sotiopoulos et al. eine klare Verbesserung der Retrieval-Qualität mit fortschreitendem Anpassungsverlauf nachweisen können, bleibt die Kombination von Einzel-Distanzen durch ein RBFN fraglich: Die durch die verborgene Neuronen-Schicht realisierte Berechnung einer radialen Ähnlichkeit von Einzel-Distanzwerten zu einem vorgegebenen Distanz-Vektor-Zentrum ist kaum zu begründen. So ist der Beitrag eines Neurons mit dem Zentrum $(0.5, 0, \dots)^T$ zur Gesamtdistanzberechnung für die Eingabedistanz-Vektoren $(0.1, 0, \dots)^T$ und $(0.9, 0, \dots)^T$ identisch; dies widerspricht klar der Intuition, dass eine (deutlich) geringere Distanz in einem Teilaspekt zu einer spürbaren Verringerung der Gesamtdistanz führen müsste.

Adaptive selbstorganisierende Karten

Während selbstorganisierende Merkmalskarten zur inhaltsbasierten Strukturierung von Dokumenten inzwischen als etabliertes Explorationsverfahren gelten, sind erst in den letzten Jahren Methoden zur Integration von Nutzer-Feedback und zur Adaption des Kartenaufbaus untersucht worden.

Erste Ansätze beschränken sich auf eine manuelle Verschiebung von platzierten Objekten [HMT07]. Die fixierende Zuweisung zu einem neuen Neuron hat hier jedoch keine Auswirkungen auf spätere Trainings-Iterationen.

Murdoch und Tzanetakis erlauben dem Benutzer die gezielte Initialisierung von Kartenzellen mit ausgewählten Modellvektoren [MT06]. So lässt sich eine Platzierung von Stücken (zumindest für die ersten Iterationen) vorgeben.

Die umfangreichsten Arbeiten zur adaptiven Kollektionsdarstellung durch selbstorganisierende Karten stammen von Nürnberger et al. Eine erste Veröffentlichung untersucht drei verschiedene Verfahren zur Anpassung an die Ordnungsstrategie des Benutzers im Kontext eines Bild-Informationssystems [NK02].

Nürnberger und Klose modellieren jedes Photo der betrachteten Kollektion als hochdimensionalen Merkmalsvektor, der sich aus dem Farbhistogramm des Bildes ableitet. Sie ordnen diese Objekte mit Hilfe einer Erweiterung des klassischen SOM-Trainings-Algorithmus, der eine datenabhängige Ausdehnung der Kartenfläche ermöglicht (Growing Self-Organizing Map, [Nü01, RMD02]), unter Verwendung einer hexagonalen Netztopologie an. Aufgrund der relativ kleinen Neuronenanzahl der verwendeten Karten wird einer Zelle in der Regel eine Vielzahl von Bildern zugeordnet; die SOMs von Nürnberger et al. realisieren damit eher eine Clustering-Aufgabe als eine feingranulare Positionierung der Objekte, wie dies bei emergenten Merkmalskarten möglich ist.

Der Nutzer kann die einem Neuron zugewiesenen Bilder betrachten und per Drag'n'Drop Photos in andere Zellen verschieben. Ferner hat er die Möglichkeit, Bilder stabil in einem

Neuron festzusetzen und so eine automatische Repositionierung durch folgende SOM-Trainingsiterationen zu verhindern.

Nürnberg und Klose unterscheiden zwischen drei alternativen Ansätzen zur Adaption an die vorgegebene Neugruppierung durch den Benutzer: Ein Verfahren zur globalen Neugewichtung von Merkmalen, eine lokale Anpassung von Merkmalsgewichten sowie eine Modifikation des Algorithmus' zur Best-Matching-Unit-Auswahl.

Globale Neugewichtung von Merkmalen Ziel der Adaption ist die Veränderung des SOM-Trainingsalgorithmus, so dass ein von Zelle s nach t verschobenes Bild i automatisch Neuron t zugeordnet wird (d.h. $bm_i \stackrel{!}{=} t$).

Hierzu wird die Euklidische Distanz, die zur Bestimmung der Best-Matching-Unit bm_i verwendet wird, zu einer Abstandsfunktion verallgemeinert, die die Gewichtung von einzelnen Merkmalsdimensionen erlaubt:

$$d_{\vec{w}}(\vec{x}_i, \vec{m}_j) = \sqrt{\sum_{k=1}^K (x_{ik} - m_{jk})^2 \cdot w_k}$$

Hierbei bezeichnet K die Dimensionalität der Feature-Vektoren und $\vec{w} = (w_1, \dots, w_K)^T$ den Gewichtsvektor.

Zur Bestimmung von $\vec{w}$ greifen Nürnberg und Klose auf folgende Heuristik zurück: Merkmalsdimensionen, in denen das Bild $\vec{x}_i$ eine geringere Abweichung von $\vec{m}_t$ als von $\vec{m}_s$ aufweist, sind Dimensionen, die die Verschiebung des Benutzers von s nach t unterstützen. Die Bedeutung der Differenz in dieser Dimension für die Best-Matching-Unit-Bestimmung ist daher zu erhöhen. Analog sind Dimensionen schwächer zu gewichten, hinsichtlich deren i näher am ungewollten s als am gewollten t liegt. Diese Idee lässt sich wie folgt formalisieren:

Es sei $\vec{e}(\vec{x}_i, \vec{m}_j)$ ein K -dimensionaler Vektor, der die punktweise Distanz zwischen den Merkmalen von Bild i und dem Modellvektor der Zelle j darstellt:

$$\vec{e}(\vec{x}_i, \vec{m}_j) = (|x_{i1} - m_{j1}|, \dots, |x_{iK} - m_{jK}|)^T$$

Dann lässt sich die durch die Repositionierung von i bewirkte Anpassung des Gewichtsvektors $\vec{w}$ als normierte Differenz zwischen der Distanz von i zu s und von i zu t formulieren:

$$\Delta \vec{w} = \frac{\vec{e}(\vec{x}_i, \vec{m}_s) - \vec{e}(\vec{x}_i, \vec{m}_t)}{\|\vec{e}(\vec{x}_i, \vec{m}_s) - \vec{e}(\vec{x}_i, \vec{m}_t)\|}$$

Der aktualisierte Gewichtsvektor $\vec{w}'$ ergibt sich bei einer monoton fallenden Lernrate μ schließlich als:

$$\vec{w}' := \vec{w} \circ (\vec{1} + \mu \cdot \Delta \vec{w}),$$

wobei $\circ$ die elementweise Vektor-Multiplikation bezeichnet.

Mit Hilfe dieser Adaptionsregel iterieren Nürnberg und Klose mehrfach über alle verschobenen Objekte und passen so den globalen Gewichtsvektor $\vec{w}$ an, bis entweder alle Bilder, für die der Nutzer feste (neue) Positionen festgelegt hat, den intendierten Zellen zugeordnet werden oder eine Höchstzahl an Durchläufen überschritten ist. In letzterem Fall bricht das Verfahren ab, ohne alle Objekte korrekt platzieren zu können.

Lokale Neugewichtung von Merkmalen Nürnberger und Klose argumentieren, dass die Anpassung des globalen Gewichtsvektors $\vec{w}$ zwar die grobe Vorstellung des Benutzers zur Verteilung von Bildern erfassen könne, die Feinstruktur erfordere jedoch eine (möglicherweise ergänzende) lokalere Adaption, die die Bedeutung von einzelnen Merkmalen in bestimmten begrenzten Bereichen abschwäche oder hervorhebe.

Das Verfahren zur Anpassung lokaler Gewichtsvektoren ähnelt dem oben beschriebenen Adaptions-Algorithmus. Im Gegensatz zu einem globalen Gewichtsvektor $\vec{w}$ wird nun jedem Neuron j ein eigener Vektor $\vec{w}_j$ zugeordnet, der bei der Berechnung des Abstandes zwischen einem Merkmalsvektor $\vec{x}_i$ und dem Modellvektor der Zelle $\vec{m}_j$ verwendet wird.

Als Reaktion auf eine Verschiebung des Bildes i von Zelle s zu t passen Nürnberger und Klose die lokalen Gewichtsvektoren von Quell- und Zielneuron an ($\vec{w}_s$ und $\vec{w}_t$). Hierzu wird ähnlich zum obigen Vorgehen erneut ein Vektor zur Erfassung der normierten punktweisen Abstände zwischen $\vec{x}_i$ und $\vec{m}_j$ berechnet:

$$\vec{e}(\vec{x}_i, \vec{m}_j) = \frac{1}{\|\vec{x}_i - \vec{m}_j\|} (|x_{i1} - m_{j1}|, \dots, |x_{iK} - m_{jK}|)^T$$

Die Gewichtung von Dimensionen mit einer hohen Distanz sollte bei der ungewollten Quell-Zelle s verstärkt und beim Ziel-Neuron t abgeschwächt werden. Dies führt zu folgenden Aktualisierungsregeln für die lokalen Gewichtsvektoren:

$$\vec{w}_s' := \vec{w}_s \circ (\vec{1} + \mu \cdot \vec{e}(\vec{x}_i, \vec{m}_s))$$

$$\vec{w}_t' := \vec{w}_t \circ (\vec{1} - \mu \cdot \vec{e}(\vec{x}_i, \vec{m}_t))$$

Analog zum Verfahren der globalen Gewichtsanzpassung werden die lokalen Gewichte in einem iterativen Verfahren angepasst.

Wie von Nürnberger und Klose beschrieben kann dieses Adaptionsverfahren zu drastischen Unterschieden der Gewichtsvektoren benachbarter Neuronen führen und damit die zentrale Eigenschaft von selbstorganisierenden Karten gefährden, dass ähnliche Objekte dicht beinander liegenden Zellen zugeordnet werden.

Modifikation der Best-Matching-Unit-Auswahl Obige Verfahren greifen in die Bestimmung der Best-Matching-Unit für ein Bild durch eine Verallgemeinerung der verwendeten Distanzfunktion zu einer *gewichteten* Euklidischen Distanz ein. Die dritte in [NK02] vorgestellte Methode verzichtet dagegen bei verschobenen Objekten vollständig auf eine Abstandsberechnung und weist dem Bild fest das Ziel-Neuron zu. Durch die anschließende Anpassung der Modellvektoren dieser Ziel-Zelle und ihrer Nachbarn im Zuge des SOM-Trainings verbreitet sich der Effekt dieser Repositionierung auf die Umgebung und bewirkt somit die Repositionierung anderer Bilder.

Mit diesem Verfahren garantieren Nürnberger und Klose zwar die (durch die anderen Verfahren nicht sicherzustellende) korrekte Positionierung der verschobenen Objekte. Die feste Platzierung von Bildern auf der Karte kommt jedoch einer nicht unproblematischen benutzerdefinierten Festlegung von Modellvektoren der betreffenden Zellen gleich: Positioniert der Anwender z.B. zwei Bilder, die vom System ähnlich repräsentiert sind, auf weit voneinander entfernten Neuronen, so entstehen zwei getrennte Kartenbereiche, die die gleichen Bildcharakteristika repräsentieren. Minimale Veränderungen der Merkmalsvektoren können damit

zu einem chaotischen Wechsel der Platzierung von Bildern zwischen diesen Regionen führen. Eine gezielte Gewichtung einzelner Merkmale im Merkmals-Vektor der Objekte ist demgegenüber in der Lage, die Unterschiede zwischen beiden Objekten herauszustellen und auch für andere Regionen der Karte nutzbar zu machen.

Gewichtsbestimmung durch quadratische Optimierung Ein viertes Verfahren, das als eine Variante der globalen Merkmals-Neugewichtung verstanden werden kann, wird von Nürnberger und Stober zur Organisation einer textuellen Dokumentensammlung entwickelt [NS08] und später im Rahmen des AUCOMA-Projektes für ein adaptives Musikinformati-
onsystem angewendet [SN09].

An die Stelle der gewichteten Distanzfunktion $d_{\vec{w}}$ tritt die Wichtung einer Cosinus-Ähnlichkeit. Wie zuvor stelle $\vec{x}_i$ den Merkmalsvektor des Stückes i dar und $\vec{m}_j$ bezeichne den Modellvektor der Zelle j . Sowohl $\vec{x}_i$ als auch $\vec{m}_j$ sind längennormiert, d.h. $\|\vec{x}_i\| = \|\vec{m}_j\| = 1$. Damit lässt sich die gewichtete Cosinus-Ähnlichkeit einfach formulieren:

$$s_{\vec{w}}(\vec{x}_i, \vec{m}_j) = \sum_{k=1}^K x_{ik} \cdot w_k \cdot m_{jk}$$

Anstatt den Gewichtsvektor in einem iterativen Greedy-Verfahren heuristisch anzupassen, formulieren Nürnberger und Stober ein quadratische Optimierungsproblem, dessen Randbedingungen die vom Benutzer vorgegebene Zuordnung eines Stückes i zur Zelle t erzwingt. Konkret formulieren Sie folgende Einschränkungen:

$$\begin{aligned} s_{\vec{w}}(\vec{x}_i, \vec{m}_t) &> s_{\vec{w}}(\vec{x}_i, \vec{m}_j) \quad \forall j \neq t \\ w_k &\geq 0 \quad \forall 1 \leq k \leq K \\ \sum_{k=1}^K w_k &= 1 \end{aligned}$$

Die erste Bedingung stellt sicher, dass der Merkmalsvektor des Stückes i die größte Ähnlichkeit zum Modellvektor der Zelle t aufweist. Dies stellt eine Positionierung von i auf t sicher. Die beiden folgenden eher technischen Einschränkungen verhindern eine schwer zu interpretierende negative Wichtung einzelner Merkmalsdimensionen bzw. vermeiden sehr große Gewichte [NS08].

Als Optimierungsziel definieren Nürnberger und Stober die Minimierung des Euklidischen Abstandes zwischen dem gesuchten Gewichtsvektor $\vec{w}$ und der gleichförmigen Gewichtung $\vec{1}$. Dies bevorzugt Lösungen $\vec{w}^*$, die eine geringe Abweichung von der unpersonalisierten Merkmalskombination aufweisen.

$$\vec{w}^* = \operatorname{argmin}_{\vec{w}} \|\vec{1} - \vec{w}\| = \operatorname{argmin}_{\vec{w}} \sum_{k=1}^K (1 - w_k)^2$$

Wie Nürnberger und Stober berichten, führt die Bestimmung der Merkmalsgewichtungen durch eine geschlossene quadratische Optimierung zu deutlich stabileren Objekt-Zuordnungen. Die Konvexität der Problemformalisierung garantiert ferner das Finden eines $\vec{w}$, das alle Randbedingungen erfüllt, falls eine Lösung existiert.

Diskussion Trotz der Attraktivität eines Optimierungsverfahrens, das aufgrund der einfachen Hinzunahme weiterer Randbedingungen sehr flexibel ist, ist seine Verwendung zur Realisierung einer Nutzeradaption im *soniXplorer* aus folgenden Gründen problematisch:

- Wie die meisten der oben vorgestellten verwandten Arbeiten setzen Nürnberger et al. die Charakterisierung eines Stückes durch einen k -dimensionalen Vektor voraus, der durch die Euklidischen Distanz oder das Cosinus-Maß mit anderen Stück-Repräsentationen verglichen werden kann.

Die bewusste Integration von ausdrucksstärkeren paarweisen Stück-Distanzfunktionen in den *soniXplorer* macht die Anwendung einer Multi-Dimensionalen Skalierung zur Umwandlung in eine solche Interpretation erforderlich. Eine solche Transformation bewahrt zwar weitestgehend die Ähnlichkeitsbeziehungen zwischen Objekten; eine direkte Beziehung zwischen einzelnen Dimensionen der gefundenen Vektoren und musikalischen Ähnlichkeitsaspekten ist jedoch nicht mehr gegeben. Wie bereits in Abschnitt 4.2.1 diskutiert, verliert die Wichtung einzelner Dimensionen damit eine unmittelbare Interpretierbarkeit. Ferner erhöht sich die Gefahr einer Überanpassung des Gewichtsvektors.

- Zur genaueren Positionierung von Titeln wird im *soniXplorer* eine emergente selbstorganisierende Karte generiert, die eine wesentlich höhere Neuronenanzahl aufweist als die von Nürnberger et al. verwendeten SOMs.

Für ein platziertes Stück muss die erste aufgeführte Randbedingung für alle bis auf eine Kartenzeile instantiiert werden. Insgesamt lässt sich die Anzahl der Nebenbedingungen durch $l \cdot (M - 1) + K + 1$ beschreiben, wobei l die Anzahl durch den Benutzer positionierter Titel und M die Anzahl der Neuronen der Karte darstellt. Wie man erkennen kann, wächst die Anzahl der Nebenbedingungen linear mit M , die Berechnungskomplexität eines quadratischen Programmes steigt jedoch überlinear mit der Anzahl der Einschränkungen und verhindert somit eine effiziente Lösung des Optimierungsproblems für größere M .

- Ebenso problematisch ist der Anstieg der Berechnungskomplexität mit der Anzahl positionierter Stücke. Nürnberger et al. gehen von einer relativ kleinen Menge von Titeln aus, die vom Benutzer verschoben oder bewusst an ihrer aktuellen Position fixiert wurden. Um ein Gefühl der Vertrautheit des Benutzers mit seiner Musiklandschaft zu erzeugen, ist jedoch die weitestgehende Stabilität von Titel-Platzierungen erforderlich. Daher werden im *soniXplorer* nur neu hinzugekommene Stücke und explizit freigesetzte Titel neu positioniert - für alle anderen müssten jeweils $(M - 1)$ Randbedingungen deklariert werden. Für eine reaktive Interaktion mit dem Explorationssystem ist daher die Lösung des vorgeschlagenen quadratischen Programmes nicht durchführbar.

4.2.2 Anpassung des Ähnlichkeitsmaßes

Wie beschrieben wertet der *soniXplorer* zwei Interaktionsmöglichkeiten des Benutzers aus, um auf dessen intendiertes Strukturierungsprinzip zu schließen:

1. Die Verschiebung eines Titels auf der Karte stellt das bewegte Stück in engere Beziehung zu benachbarten Titeln.

2. Die Veränderung des Höhenprofils trennt Stück-Gruppen voneinander oder schafft ein Cluster zusammengehöriger Titel.

Zur Berücksichtigung dieser Vorgaben lässt sich in verschiedene Phasen der Prozesskette von der Distanzfunktions-Kombination über die Multidimensionale Skalierung bis hin zur Erstellung der selbstorganisierenden Karte (vergl. Abbildung 1.11) eingreifen.

Wie in Abschnitt 4.2.1 bei der Diskussion der Best-Matching-Auswahl-Modifikation in Nürnbergers Adaptionungsverfahren bereits angedeutet, birgt die isolierte Modifikation der Modell-Vektoren von Kartenregionen die Gefahr einer chaotischen Positionierung von neuen oder freigesetzten Stücken zwischen Bereichen ähnlicher Modell-Vektoren.

Sinnvoller erscheint es, die Basis der Stück-Gruppierung auf der Karte, nämlich das zugrunde liegende Ähnlichkeitsmaß selbst, den Vorgaben des Benutzers anzupassen und so sein individuelles Ähnlichkeitsverständnis zu erlernen. Diese Adaption des Ähnlichkeitsmaßes wird im *soniXplorer* kombiniert mit der Fixierung nicht beweglicher Stücke auf ihrer aktuellen Position. Hierdurch wird eine Neuverteilung der etablierten Gruppenstrukturen auf der Karte verhindert und die Stabilität der groben Kartenstruktur gewährleistet.

Wie in Abschnitt 2.7 diskutiert, erbringt die Ersetzung eines linearen Modells zur Verknüpfung von Einzel-Distanzen durch eine komplexere Support-Vector-Regressor keinen deutlichen Qualitätsgewinn. Daher beschränkt sich der *soniXplorer* auf die Anpassung der Gewichtung der implementierten Ähnlichkeits-Teilaspekte.

Aufbau einer Ziel-Distanz-Matrix

Aus den Karten-Änderungen des Benutzers wird hierzu zunächst eine Ziel-Distanz-Matrix T abgeleitet. Da die Distanzen in T als symmetrisch angenommen werden, genügt die Erfassung der oberen Dreiecksmatrix.

Grundlage für die Berechnung der vom Anwender intendierten Distanz zwischen Stück i und j ist deren Position $\vec{p}_i$ bzw. $\vec{p}_j$ auf der Karte. Durch bilineare Interpolation lässt sich einer Position $\vec{p}$ ein Modell-Vektor-Wert zuweisen, auch wenn $\vec{p}$ zwischen den Gitterpositionen der Neuronen auf der Karte liegt.

Sei hierzu $\vec{p} = (x, y)$, d.h. x und y sind die zweidimensionalen Landschafts-Koordinaten von $\vec{p}$. $\vec{m}_{\mathbb{N}}(x', y')$ für $x', y' \in \mathbb{N}$ bezeichne den Modellvektor für das Neuron an Position (x', y') . Damit lässt sich der Modell-Vektor an Position $\vec{p}$ wie folgt interpolieren:

$$\begin{aligned}\vec{m}_u(\vec{p}) &= (\lceil x \rceil - x) \cdot \vec{m}_{\mathbb{N}}(\lfloor x \rfloor, \lfloor y \rfloor) + (x - \lfloor x \rfloor) \cdot \vec{m}_{\mathbb{N}}(\lceil x \rceil, \lfloor y \rfloor) \\ \vec{m}_d(\vec{p}) &= (\lceil x \rceil - x) \cdot \vec{m}_{\mathbb{N}}(\lfloor x \rfloor, \lceil y \rceil) + (x - \lfloor x \rfloor) \cdot \vec{m}_{\mathbb{N}}(\lceil x \rceil, \lceil y \rceil) \\ \vec{m}(\vec{p}) &= (\lceil y \rceil - y) \cdot \vec{m}_u(\vec{p}) + (y - \lfloor y \rfloor) \cdot \vec{m}_d(\vec{p})\end{aligned}$$

Hieraus ergibt sich ein erster Ansatz für die Belegung des Ziel-Distanzmatrix-Eintrags T_{ij} , der die vom Benutzer intendierte Distanz zwischen Stück i und j quantifiziert:

$$T_{ij}^{dist} = \|\vec{m}(\vec{p}_i) - \vec{m}(\vec{p}_j)\|$$

In ersten Experimenten erwies sich die Verwendung dieser Modell-Vektor-Distanz als informativer gegenüber der zweidimensionalen Koordinaten-Distanz $\|\vec{p}_i - \vec{p}_j\|$.

Die Veränderung des Terrains durch den Benutzer sollte zu einer stärkeren Trennung (Abstandsvergrößerung) oder zum Abbau von separierenden Höhenzügen (Abstandsverringering) zwischen zwei Stücken führen. Zur Integration dieser Veränderung in die Ziel-Distanz-Matrix wird das veränderte Höhenprofil auf der direkten Verbindung zwischen $\vec{p}_i$ und $\vec{p}_j$ mit seinem Zustand vor der Modifikation verglichen:

$$T_{ij}^{height'} = \frac{1}{\|\vec{p}_i - \vec{p}_j\|} \cdot \left(\int_0^1 h_{new}((1-s)\vec{p}_i + s \cdot \vec{p}_j) ds - \int_0^1 h_{old}((1-s)\vec{p}_i + s \cdot \vec{p}_j) ds \right),$$

wobei $h_{new}(\vec{p})$ den modifizierten Höhenwert der Landschaft im Punkte $\vec{p}$ und $h_{old}(\vec{p})$ den Höhenwert vor einer Veränderung durch den Benutzer quantifiziert.

Obiges Integral wird durch Diskretisierung der Strecke zwischen $\vec{p}_i$ und $\vec{p}_j$ bei einer Schrittweite von sw approximiert:

$$T_{ij}^{height} = \frac{1}{\lfloor \frac{\|\vec{p}_i - \vec{p}_j\|}{sw} \rfloor + 1} \cdot \sum_{s=0}^{\lfloor \frac{\|\vec{p}_i - \vec{p}_j\|}{sw} \rfloor} \left(h_{new}((1-s \cdot sw)\vec{p}_i + s \cdot sw \cdot \vec{p}_j) - h_{old}((1-s \cdot sw)\vec{p}_i + s \cdot sw \cdot \vec{p}_j) \right)$$

Während der Arbeit mit dem **soniXplorer** wurde eine Schrittweite von $sw = 0.1$ empirisch als hinreichend fein identifiziert.

Für einen Parameter μ , der den Einfluss der Höhenveränderung auf die intendierten Distanzen bestimmt, ergibt sich der Zielmatrix-Eintrag T_{ij} abschließend wie folgt:

$$T_{ij} = T_{ij}^{dist} + \mu \cdot T_{ij}^{height}$$

Multiple Lineare Regression

Die Aufgabe des Adaptionprozesses besteht nun darin, die implementierten Distanzfunktionen so zu kombinieren, dass die gewichteten Abstände zwischen zwei Stücken i und j dem Matrix-Eintrag T_{ij} gleichen:

$$T_{ij} \stackrel{!}{=} \sum_{k=1}^K w_k \cdot d_k(i, j) + w_0,$$

wobei K die Anzahl verfügbarer Einzeldistanzen beschreibt und $d_k(i, j)$ die Distanz zwischen Stück i und j bzgl. der k -ten solchen Distanz quantifiziert. Gesucht sind (für alle i, j gültige) Gewichte w_k sowie der konstante Offset w_0 .

Instantiiert man für jedes (nicht-triviale) Paar von N' fixierten Stücken obige Bedingung, erhält man ein lineares System aus $L = \binom{N'}{2}$ Gleichungen:

$$\begin{pmatrix} T_{1,2} \\ T_{1,3} \\ \vdots \\ T_{N'-1,N'} \end{pmatrix} = \begin{pmatrix} 1 & d_1(1,2) & d_2(1,2) & \cdots & d_K(1,2) \\ 1 & d_1(1,3) & d_2(1,3) & \cdots & d_K(1,3) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & d_1(N'-1,N') & d_2(N'-1,N') & \cdots & d_K(N'-1,N') \end{pmatrix} \cdot \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_K \end{pmatrix}$$

Für die folgende Darstellung sei dieses Gleichungssystem kurz mit $Y = X \cdot \vec{w}$ bezeichnet.

Da die Anzahl der Gleichungen L in der Regel deutlich die Anzahl der gesuchten Parameter $K + 1$ im Gewichtsvektor $\vec{w}$ übersteigt, ist das System überbestimmt. Anstatt einer exakten

Auflösung nach $\vec{w}$ wird daher ein Optimierungsproblem formuliert: Gesucht sind Gewichte w_i , so dass die Summe der quadratischen Abweichungen zwischen linker und rechter Seite der Einzelgleichungen minimal wird. Formal ausgedrückt:

$$\vec{w}^* = \operatorname{argmin}_{\vec{w}} \sum_{l=1}^L (y_l - \vec{x}_l \cdot \vec{w})^2$$

Damit ist das Problem als Multiple Regressions-Aufgabe formuliert und kann geschlossen gelöst werden (vergl. z.B. [DHS01]):

$$\vec{w}^* = (X^T X)^{-1} X^T Y$$

Zur Berechnung dieser Lösung greift der **soniXplorer** auf eine Matlab-Implementierung zurück, die die aufwändige Matrix-Multiplikation und -Inversion zu Gunsten eines effizienteren und numerisch stabileren Verfahrens ersetzt. Versuche mit verschiedenen Problemgrößen zeigten ein praktisch lineares Laufzeit-Wachstum in L und eine Berechnungszeit von unter 1s für $L = 1\,000\,000$ auf einem Pentium M mit ca. 2 GHz.

Fehlende Distanzwerte

Die Kombination von Einzeldistanzen zu einem zusammenfassenden Abstandswert setzt deren Berechenbarkeit voraus. Wie in Abschnitt 2.8.4 ausgeführt, ist diese Bedingung für kollaborative Ähnlichkeitsmaße nicht notwendigerweise erfüllt, da das Stück z.B. aufgrund seiner Audio-Daten nicht identifiziert und somit im Internet-Portal lokalisiert werden konnte oder (noch) nicht hinreichend viele Nutzer-Informationen für eine tragfähige Ähnlichkeitsbestimmung vorliegen.

Diesem Problem fehlender Einzeldistanzwerte wird im **soniXplorer** wie folgt begegnet: Für jede Teilmenge kollaborativer Ähnlichkeiten wird ein separates lineares Modell gelernt, deren Trainingsmenge aus allen Paaren fixierter Stücke gebildet wird, für die (mindestens) alle betrachteten Distanzwerte vorliegen.

Die Induktion eigener Modelle ermöglicht später während der Berechnung einer Gesamtdistanz zwischen zwei ungesetzten Stücken i und j die gezielte Auswahl eines Gewichtsvektors, der auf Grundlage einer maximal großen Trainings-Datenbasis ermittelt wurde.

Da im **soniXplorer** Abstände zwischen Stücken in Form von Distanzmatrizen gespeichert werden, ist nicht nur der Speicheraufwand für diese Matrizen quadratisch in der Anzahl betrachteter Stücke; gleiches gilt naturgemäß auch für die Anzahl der Lerninstanzen ($L \in \mathcal{O}(N^2)$).

Eine Möglichkeit zur Reduktion dieser Komplexität ist die Beschränkung der Ähnlichkeitsberechnung für ein Stück i auf Distanzen zu ausgewählten Stücken („Anker-Stücke“), die eine besondere Repräsentativität für den Benutzer haben. Auch wenn diese Funktionalität in der bisherigen Version des **soniXplorer** noch nicht endgültig implementiert ist, wurden die Verfahren zur Gewichtsvektor-Adaption und zur Transformation in eine Vektor-Repräsentation (siehe Abschnitt 4.2.3) bereits für diese Erweiterung ausgelegt.

So übergeht obige Auswahlprozedur für Trainingsdatensätze schlicht fehlende Abstandswerte zwischen den Stücke i und j bei der Induktion aller linearen Modelle.

Distanzmatrix-Zusammenstellung

Um ein Stück in die Explorationsumgebung neu aufzunehmen oder ein freigesetztes zu repositionieren, ist die Berechnung von Distanzen zu (in der Regel allen) anderen Titeln erforderlich. Zur Ermittlung eines Abstandes zwischen Stück i und j wird zunächst das lineare Modell identifiziert, das maximal viele verfügbare kollaborative Einzeldistanzen berücksichtigt. Für die folgende Darstellung kombiniere dieses Modell $\tilde{K}$ Teilaspekte und gewichte die Einzelabstände $d_{in(1)}(i, j), \dots, d_{in(\tilde{K})}(i, j)$ durch den Vektor $(\tilde{w}_0, \dots, \tilde{w}_{\tilde{K}})^T$.

Der Abstand zwischen fixierten Stücken ist durch die Vorgabe des Benutzers festgelegt, die sich im Eintrag in der Ziel-Distanz-Matrix T niederschlägt. Im Idealfall einer perfekten Adaption entspricht dieser der Einzelmaß-Kombination.

Für die Distanzmatrix D , die im weiteren Prozess zur Quantifizierung des Abstandes zwischen Stücken herangezogen wird, lässt sich somit folgende Berechnungsformel aufstellen:

$$D_{ij} = \begin{cases} T_{ij} & \text{falls } i \text{ und } j \text{ fixiert} \\ \sum_{k=1}^{\tilde{K}} \tilde{w}_k \cdot d_{in(k)}(i, j) + \tilde{w}_0 & \text{sonst} \end{cases}$$

Gemäß obiger Bemerkungen zur Reduktion des Ähnlichkeitsberechnungsaufwandes kann diese Matrix fehlende Werte aufweisen, falls Distanzen zwischen Stücken (bewusst) nicht betrachtet werden.

4.2.3 Anpassung der Vektor-Repräsentation

Wie in der Prozessskizze in Abbildung 1.11 zu erkennen ist, folgt der Berechnung paarweiser Distanzen die Zuweisung einer Vektorrepräsentation zu Stücken unter weitest möglicher Beibehaltung der vorgegebenen Abstände zwischen den Titeln.

In Abschnitt 3.2 wurden für diese Aufgabe bereits verschiedene alternative Lösungsstrategien untersucht und eine iterative Implementierung der Multidimensionalen Skalierung ausgewählt, die den aufsummierten gewichteten Abstand zwischen paarweiser Distanz und Euklidischem Abstand der Vektorrepräsentation (*weighted Stress*) minimiert:

$$\text{Stress}_{\text{weighted}}(\vec{x}_1, \dots, \vec{x}_N) = \sqrt{\frac{\sum_{\substack{i,j=1\dots N \\ i \neq j}} r_{ij} (D_{ij} - \|\vec{x}_i - \vec{x}_j\|)^2}{\sum_{\substack{i,j=1\dots N \\ i \neq j}} r_{ij} D_{ij}^2}}$$

Hierbei bezeichne r_{ij} das Gewicht (oder die „Relevanz“) der Abweichung zwischen paarweiser Distanz D_{ij} und Euklidischem Abstand der zugewiesenen Vektoren $\vec{x}_i$ und $\vec{x}_j$ für die Gesamtqualität der Lösung.

Diese Erweiterung des klassischen Stress-Kriteriums ermöglicht den direkten Umgang mit fehlenden Distanzwert-Einträgen: Abstände, die aus Effizienz-Gründen bewusst nicht ermittelt wurden, erhalten eine Gewichtung von 0 und haben somit keinerlei Einfluss auf die Positionierung der Stücke im (hochdimensionalen MDS-) Raum.

In einer späteren Ausbaustufe des soniXplorers sind auch Werte zwischen 0 und 1 für die Gewichtung r_{ij} denkbar, die zur Quantifizierung des Grades der Fixierung eines Stückes genutzt

werden könnten. Dies würde dem Benutzer z.B. ermöglichen, die ungefähre Position eines Stückes auf der Karte festzulegen, die exakte Positionierung jedoch dem System zu überlassen. Eine Implementierung und Evaluation dieser Idee bleibt jedoch zukünftigen Arbeiten vorbehalten.

Um die Positionierung fixierter Stücke im Rahmen der SOM-Generierung zu vereinfachen, ist es wünschenswert, wenn bereits die Multidimensionale Skalierung die Vektorrepräsentation festgesetzter Titel so wenig wie möglich verändert. Dies lässt sich unter Beibehaltung des eigentlichen Optimierungsziels der $\text{Stress}_{\text{weighted}}$ -Minimierung durch Initialisierung des iterativen MDS-Verfahren mit geeigneten Startbedingungen erreichen.

Für ein fixiertes Stück i wird der interpolierte Modellvektor $\vec{m}(p_i)$ seiner Position p_i als Anfangsvektor verwendet. Die Vektorrepräsentation freigesetzter oder neuer Titel wird mit einer gewichteten Mittelwert aus den Vektoren umgebender fester Stücke initialisiert.

Hierzu sei $Neigh(i)$ die Menge der fixierten Stücke in der Umgebung eines ungesetzten Titels i . Im *soniXplorer* werden die ersten $maxNeigh = 5$ Stücke aus einer aufsteigend nach Distanz zu i sortierten Titel-Liste verwendet. Der Einfluss eines Stückes $j \in Neigh(i)$ auf den Vektor-Mittelwert ist proportional zum Abstand D_{ij} . Somit errechnet sich der gesuchte Anfangsvektor für i wie folgt:

$$\vec{init}_i = \sum_{j \in Neigh(i)} \vec{m}(p_j) \cdot \left(\frac{1}{\sum_{j' \in Neigh(i)} \frac{1}{D_{ij'}}} \cdot \frac{1}{D_{ij}} \right)$$

4.2.4 Anpassung der selbstorganisierenden Karte

Als letzter Schritt bleibt die Anpassung der selbstorganisierenden Karte (Abbildung 1.11).

Um die gewonnene Vertrautheit des Benutzers mit der Umgebung, die er möglicherweise durch Hinweisschilder, bewusste Stückrepositionierungen und Terrain-Veränderungen angepasst hat, nicht zu gefährden, kommt der Stabilität der Platzierung von nicht freigesetzten Stücken ein großer Stellenwert zu.

Daher fixiert der *soniXplorer* alle Titel, die nicht explizit vom Benutzer zur Repositionierung freigegeben wurden, exakt an der Ursprungsposition. Hierzu wird analog zum Ansatz von Nürnberger (Abschnitt 4.2.1, „Modifikation der Best-Matching-Unit-Auswahl“) in den Prozess der Auswahl der Best-Matching-Unit eines Stückes eingegriffen: An die Stelle der Suche nach dem Neuron mit dem geringsten Modellvektor-Abstand zum Titel tritt die konstante Auswahl der Zelle, die der aktuellen Position des Stückes auf der Karte am nächsten ist.

Dieses Vorgehen garantiert nicht nur die Stabilität der Stück-Platzierung. Die Angleichung der Modellvektoren in der Umgebung eines fixierten Stückes stellt ferner sicher, dass sich die Modellvektor-Struktur der Landschaft ihrerseits an die forcierte Positionierung anpasst und somit die sinnvolle (und insbesondere Topographie-erhaltende) Platzierung neuer oder freigesetzter Titel gewährleistet.

Kapitel 5

Evaluation

Zielsetzung dieser Arbeit ist die Entwicklung und Evaluation eines Prototypen zur multi-modalen adaptiven Exploration von Musikkollektionen, der dem Benutzer eine bessere Orientierung über seine Stücksammlung verschafft und hierdurch einen schnelleren Zugriff auf gesuchte Titel ermöglicht, auch wenn das Informationsbedürfnis nicht klar artikuliert werden kann.

Um zu ermitteln, inwieweit diese Anforderungen mit dem entwickelten Prototypen *soniXplorer* erfüllt werden können, wurde im Frühjahr und Sommer 2009 eine Evaluation des Systems mit insgesamt neun Probanden durchgeführt. Die folgenden Abschnitte dieses Kapitels beschreiben und begründen Aufbau und Ergebnisse der Untersuchungen.

5.1 Fragestellungen und Untersuchungshypothesen

Zentrale Voraussetzung für eine positive Explorationserfahrung des Benutzers ist die Schaffung einer Umgebung, in der sich der Anwender intuitiv orientieren und bewegen kann.

Neben der detaillierten Erfassung des subjektiven Eindrucks der Probanden zur Explorationsumgebung wird im Rahmen dieser Evaluation daher insbesondere der Frage nachgegangen, inwiefern einzelne Komponenten des *soniXplorers* zur Verbesserung der *Orientierung* des Anwenders in der betrachteten Kollektion beitragen.

Konkret soll die Gültigkeit der folgenden Arbeitshypothesen überprüft werden:

- *Orientierungsgewinn durch SOM-Strukturierung*

Die Platzierung von Titeln auf einer zweidimensionalen Landkarte mit der Vorgabe, dass nach einem generischen Modell als perzeptuell ähnlich betrachtete Stücke in benachbarten Regionen positioniert und von Bergketten getrennt sind, verbessert die Orientierung des Benutzers in der Landschaft und ermöglicht eine gezieltere Navigation in der gegebenen Musikkollektion.

- *Orientierungsgewinn durch akustischen Explorationsraum*

Die entwickelte Ergänzung der visuellen Darstellung durch eine permanente direktionale akustische Wiedergabe von Stücken in der unmittelbaren Umgebung des Benutzers ermöglicht einen intuitiveren Zugriff auf die Kollektion und verbessert so die Orientierung des Anwenders in der Landschaft.

- *Anpassungsfähigkeit an unterschiedliche Ähnlichkeitsverständnisse*

Das implementierte Verfahren zur Adaption des zugrunde liegenden Distanzmaßes an das intendierte Ähnlichkeitsverständnis des Benutzers auf der Basis seiner Interaktion mit dem System ermöglicht eine personalisierte Modellierung unterschiedlicher Strukturierungsprinzipien.

5.2 Evaluations-Design

Um die Gültigkeit der obigen Hypothesen objektiv bewerten zu können, werden in Szenario-basierten Versuchen quantitative Messwerte ermittelt (Abschnitt 5.2.1). Für die Qualität eines Informationssystems ist darüber hinaus jedoch ebenso die Akzeptanz der Applikation beim Nutzer ausschlaggebend [JOPP98]. Daher wird die quantitative Auswertung durch subjektive Bewertungen ergänzt, die mit Hilfe eines umfangreichen Fragebogens erfasst werden (Abschnitt 5.2.2).

Zunächst erhalten die Probanden eine Einführung in das Explorationssystem anhand einer unabhängigen im weiteren Evaluationsverlauf nicht mehr verwendeten Kollektion. Sie werden mit der Navigation in der virtuellen Umgebung mit Hilfe des Game-Controllers und der Eingrenzung des akustischen Wahrnehmungswinkels zur Fokussierung auf frontal positionierte Stücke vertraut gemacht. Darüber hinaus werden Möglichkeiten zur Anpassung der Landschaft wie das Setzen und Beschriften von Hinweisschildern, Veränderungen des Terrain-Höhenprofils und Verschiebungen der platzierten Stücke erläutert. Für diese Phase des Kennenlernens des Systems ist kein Zeitlimit vorgegeben.

5.2.1 Quantitative Auswertung

Ziel der ersten Versuche ist eine quantitative Bewertung der Unterstützung, die der *soniXplorer* dem Benutzer zur Orientierung in einer Musikkollektion bietet. Hierzu wird ein Szenario simuliert, in dem der Anwender ein Stück identifizieren möchte, von dem er eine klangliche Vorstellung hat, dessen beschreibende Metadaten (wie z.B. Titelbezeichnung, Künstlername oder CD-Cover-Bild) ihm jedoch nicht vorliegen. Dies entspricht einem typischen Fall explorativer Suche, bei der das Informationsbedürfnis nicht klar artikulierbar ist.

Als Messgröße dient die Zeit, die der Benutzer (bei Aktivierung oder Deaktivierung unterschiedlicher Funktionalitäten des *soniXplorers*) zur Erfüllung dieser Suchaufgabe in der Explorationsumgebung benötigt. Um die Ergebnisse der verschiedenen Versuchsteilnehmer vergleichen zu können, werden einheitliche Musikkollektionen und zu lokalisierende Stücke verwendet.

Orientierungsgewinn durch SOM-Strukturierung

Im ersten Experiment wird die Relevanz der Strukturierung der Kollektion durch eine selbstorganisierende Merkmalskarte und deren Visualisierung und Auralisierung für die Orientierung des Benutzers in der Explorationsumgebung untersucht.

Hierzu werden für jeden Benutzer zwei Zeitmessungen gegenübergestellt: Im ersten Ansatz muss der Anwender ein akustisch präsentiertes Stück auf einer Karte lokalisieren, auf der die

Titel der Kollektion zufällig verteilt sind. Daran anschließend wird die gleiche Aufgabe für ein neues Stück gestellt. Dieses Mal sind die Titel der Musiksammlung jedoch gemäß dem (nicht personalisierten) initialen Ähnlichkeitsmaß mit Hilfe einer SOM gruppiert. Tabelle A.2 im Anhang beschreibt die Zusammensetzung der verwendeten Stückkollektion. Zu lokalisieren sind die Titel 492 (Hillard Ensemble: *Jesu Deine Passion*) und 311 (Edvard Grieg: *From Holberg's Time - Preludium*).

Vor der Präsentation des zu suchenden Stückes hat der Proband jeweils fünf Minuten Zeit, um sich mit der Struktur der Karte vertraut zu machen und z.B. Hinweisschilder zur Kennzeichnung von Regionen der Landschaft zu setzen. Da in diesem Versuch die Qualität der generierten Stück-Positionierung untersucht werden soll, ist eine Repositionierung der Titel auf der Karte nicht erlaubt.

Um eine systematische Abhängigkeit der benötigten Zeit vom zu suchenden Stück auszuschließen, wird bei der Hälfte der Versuchsteilnehmer das zweite Stücke mit der unstrukturierten Karte (und das erste mit der strukturierten SOM) kombiniert.

Ebenso wird die Reihenfolge der sich so ergebenden Aufgaben variiert, um einen systematischen Effekt der möglicherweise einsetzenden wachsenden Vertrautheit der Versuchsteilnehmer mit der gleichbleibenden Musikkollektion auf die gemessene Suchzeit zu vermeiden. So wird die Hälfte der Versuchgruppe zunächst mit der strukturierten Landschaft und erst im Anschluss daran mit der zufälligen Stück-Verteilung konfrontiert.

Orientierungsgewinn durch akustischen Explorationsraum

Die überwiegende Mehrzahl der bekannten Explorationsumgebungen für Musikkollektionen beschränkt sich auf eine visuelle Präsentation strukturiert angeordneter Stücke. Um einen dargestellten Titel akustisch wahrzunehmen, muss der Benutzer explizit ähnlich einem Standard-Medienplayer die Wiedergabe initiieren. Eine akustische Orientierung des Anwenders in der Umgebung durch permanent erklingende Schallquellen ist zumeist nicht vorgesehen.

Der soniXplorer dagegen kombiniert visuelle und akustische Darstellung zu einer multimodalen Explorationsumgebung. In einem zweiten Experiment mit der Versuchsgruppe wird die Bedeutung dieser Erweiterung der Visualisierung um eine permanente Auralisierung untersucht.

Der Versuchsaufbau ähnelt dem des ersten Experiments: Wiederum wird die Verkürzung der Suchzeit nach einem vorgegebenen Stück bei unterschiedlicher Funktionalität der Explorationsumgebung als Indiz für die Bedeutung dieser Funktionalität ausgewertet.

Die Tabellen A.3 und A.4 führen die Titel der verwendeten beiden unabhängigen Musikkollektionen auf. In der ersten Stückmenge ist der Country-Song 551 (Billie Jo Spears: *Blanket On the Ground*) zu lokalisieren; in der zweiten Titelsammlung muss ein Satz aus dem *Menuett d-Moll* von Johann Sebastian Bach für Cembalo und Flöte (ID 324) in der Explorationslandschaft gefunden werden.

In beiden Fällen werden die Suchstücke wiederum rein akustisch präsentiert. Beide Kollektionen werden anhand des initialen generischen Ähnlichkeitsbegriffes strukturiert und als Landschaft visualisiert.

Im ersten Ansatz steht dem Anwender lediglich der SOLO-Modus zur akustischen Wiedergabe zu Verfügung: Zum Abspielen eines Titels muss dieser bewusst ausgewählt und die Wiedergabe explizit gestartet werden.

Für die Erfüllung der zweiten Suchaufgabe dagegen kann der Benutzer sich zusätzlich anhand der ihn umgebenden akustischen Aura orientieren, die vom *soniXplorer* unter Berücksichtigung der benutzergesteuerten fokussierten Wiedergabe frontaler Titel generiert wird.

Wie im ersten Evaluationsexperiment werden Aufgabenstellung und Funktionalitätsreduktion für die Hälfte der Versuchgruppe variiert, um einen systematischen Effekt der gewählten Stückmenge auf die Veränderung der Suchzeit auszuschließen.

Anpassungsfähigkeit an unterschiedliche Ähnlichkeitsverständnisse

Ein weiteres Merkmal, das den *soniXplorer* von anderen Explorationsumgebungen für Musikkollektionen differenziert, ist seine Fähigkeit zur Adaption an einen individuellen Ähnlichkeitsbegriff des Benutzers, indem die Kombination von Distanzmaßen, die einzelne musikalische Aspekte modellieren, so angepasst wird, dass sie die Veränderungen der Landschaft und Stück-Platzierungen durch den Anwender bestmöglich erklärt.

Zur Bewertung der Leistungsfähigkeit dieses Ansatzes wird die Versuchsguppe mit einer überschaubaren Kollektion von 15 Stücken konfrontiert (eine detaillierte Auflistung der Titel findet sich in Tabelle A.5). Diese Sammlung wird zunächst gemäß dem unpersonalisierten initialen Ähnlichkeitsbegriff mit Hilfe einer SOM aus $20 \cdot 20$ Zellen strukturiert und als dreidimensionale Landschaft visualisiert und auralisiert.

In einem ersten Versuch werden die Probanden aufgefordert, die Gruppierung der Stücke durch Repositionierung einzelner Titel ihrem persönlichen Ähnlichkeitsempfinden anzupassen und die dargestellte Landschaft ggf. durch die Ausbildung oder Abflachung von trennenden Bergketten zu verändern.

Die Fähigkeit des *soniXplorers* zur Adaption dieses so artikulierten Ähnlichkeitsbegriffes wird mit Hilfe eines Verfahrens quantifiziert, das an die Leave-One-Out-Methodik zur Klassifikator-Bewertung erinnert [Mit97]:

Jedes Stück wird einzeln aus der Kollektion entfernt und als neu hinzukommender Titel betrachtet, für den auf der Grundlage des unangepassten Ähnlichkeitsbegriffes eine neue Position im Kontext der angepassten Karte bestimmt wird. Ein Vergleich dieser Position mit der ursprünglich vom Benutzer vorgegebenen Platzierung resultiert in dem Positionierungsfehler e_u , der sich als Distanz der Kartenkoordinaten berechnet.

Anschließend wird aus der Platzierung der verbleibenden 14 Stücke eine Neugewichtung der Einzel-Distanzmaße ermittelt. Nach dieser Adaption wird das betrachtete Stück erneut automatisch auf der Karte platziert. Wieder werden die Koordinaten der zugewiesenen Position mit der Benutzervorgabe verglichen und als Fehler e_a quantifiziert.

Alternativ zur Distanz der Koordinaten wurden für e_u und e_a die Distanz der dortigen (interpolierten) Modellvektoren betrachtet. Da diese jedoch keine deutlichen Unterschiede zum Koordinaten-Abstand aufwies, wird im folgenden auf eine weitere Diskussion der Modellvektoren-Distanz verzichtet.

Nach Iterierung dieses Verfahrens für alle 15 Stücke der Kollektion werden die Einzelfehler durch Mittelwertbildung pro Proband zusammengefasst.

Der Vergleich des Positionierungsfehlers bei unangepasstem (e_u) und angepasstem (e_a) Distanzmaß kann als Indiz für die Anpassungsfähigkeit des Systems an das subjektive Ähnlichkeitsverständnis des Benutzers interpretiert werden.

Eine starke Verringerung des Positionierungsfehlers ließe sich jedoch auch durch eine schlechte initiale Kombination der Einzeldistanzen erklären und ist noch kein Beleg für die Anpassungsfähigkeit an *verschiedene* Ähnlichkeitsverständnisse. Daher werden die Probanden in einem weiteren Versuch gebeten, an die Stelle ihres persönlichen Strukturierungsprinzips ein vorgegebenes Ordnungskriterium zu setzen: Zur Simulation der Organisation einer Musikkollektion durch einen DJ sollen die 15 Titel gemäß ihres *Tempos* angeordnet werden.

Wiederum wird der Unterschied des Positionierungsfehlers unter dem aktuellen subjektiven Distanzmaß zum Positionierungsfehler nach einem Anpassungsschritt ermittelt. Eine signifikante Verringerung deutet auf die Fähigkeit des *soniXplorers* hin, unterschiedliche Ordnungsprinzipien durch geeignete Neugewichtung der Einzeldistanzmaße zu adaptieren.

5.2.2 Subjektive Auswertung

Mit Hilfe eines umfangreichen Fragebogens wird im Anschluss an die oben beschriebenen quantitativen Experimente ein subjektiver Eindruck der Probanden vom System ermittelt.

Jeder der Versuchsteilnehmer wird mit insgesamt 26 Aussagen aus fünf Themengruppen konfrontiert und kann den Grad seiner Zustimmung in sieben Stufen von „strongly agree“ (1) bis „strongly disagree“ (7) angeben oder auf eine Stellungnahme verzichten. Darüber hinaus wird Platz für freie Kommentare und Verbesserungsvorstellungen eingeräumt. Eine vollständige Zusammenstellung des verwendeten Fragebogens findet sich in Anhang B.

Die fünf Themenegruppen behandeln zum einen die zentralen Kern-Merkmale des *soniXplorers* und erfragen darüber hinaus die empfundene Benutzerfreundlichkeit des Systems:

- *Strukturierte Kollektionsdarstellung*
Wie hilfreich wird die Strukturierung der Umgebung zur Suche nach Stücken empfunden? Wie hilfreich ist die Möglichkeit zur Umformung des Terrains? Wie hilfreich und notwendig ist die Möglichkeit, sich durch eine Erhöhung der Blickposition einen Überblick über die Landschaft zu verschaffen?
- *Visuelle Darstellung*
Wie gefällig und intuitiv ist die grafische Präsentation des Terrains? Wie wird die Bedeutung der Darstellung von Alben-Texturen eingeschätzt? Wie intuitiv ist die visuelle Charakterisierung unterschiedlicher Eigenschaften von Stücken (wie z.B. Markierung, Rotation bei akustischer Wiedergabe, etc.)?
- *Akustische Darstellung*
Wie hilfreich oder verwirrend ist die simultane akustische Wiedergabe von Titeln in der Umgebung? Wie hilfreich ist die Möglichkeit zur Einschränkung des Wahrnehmungswinkels? Wie hoch ist die Bedeutung des Solo-Modus? Wie wichtig ist die Funktionalität zum expliziten Ausschluss bzw. zur expliziten Aufnahme von Stücken in die Wiedergabemenge? Ist die Anzahl gleichzeitig erklingender Stücke zu hoch? Wie intuitiv verständlich ist das Verfahren zur Selektion gleichzeitig wiederzugebender Titel? Wie nützlich und intuitiv ist die implementierte Abschattung von Klangquellen, die sich hinter einer Berghöhe befinden?
- *Hinweisschilder*
Wie nützlich und hilfreich ist die Möglichkeit zur Platzierung von Hinweisschildern? Ist die implementierte Funktionalität zur Anpassung von Hinweisschildern ausreichend?

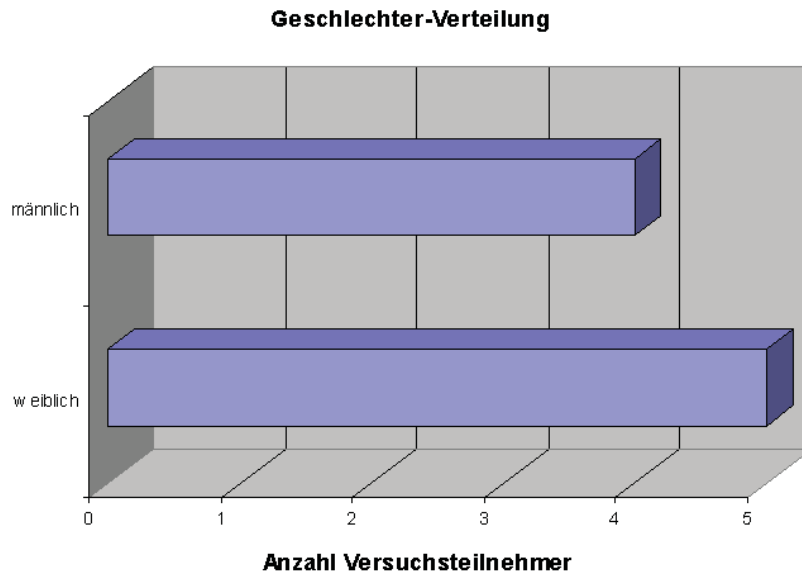


Abbildung 5.1: Geschlechter-Verteilung in der Versuchsgruppe

- *Allgemeine Benutzerfreundlichkeit*

Wie intuitiv ist die Navigation auf dem Grund mit Hilfe des Game-Controllers? Wie einfach und intuitiv ist die Veränderung der Betrachterhöhe? Wie einfach und intuitiv ist die Selektion von Titeln? Wie bedienbar ist die Funktionalität zur Eingrenzung des Wahrnehmungswinkels? Wie intuitiv lassen sich Stücke in der Landschaft verschieben? Wie benutzerfreundlich sind die Erstellung, das Löschen und das Anpassen von Hinweisschildern? Wie intuitiv ist die Umformung des Terrains durchführbar? Wie motivierend und unterhaltsam ist die Exploration einer Musikkollektion mit dem soniXplorer?

5.3 Ergebnisse und Auswertung

Dieser Abschnitt führt nach einer Beschreibung der Versuchsgruppe die Resultate der durchgeführten Experimente auf und diskutiert die Auswertung der verteilten Fragebögen.

5.3.1 Versuchsgruppe

Die im folgenden beschriebenen Versuche wurden im Frühjahr und Sommer 2009 mit insgesamt neun Probanden aus dem Umfeld des Autors in Muscat (Oman) und Aachen durchgeführt.

Mit vier männlichen und fünf weiblichen Teilnehmern weist die Gruppe eine ausgeglichene Geschlechter-Verteilung auf (vergl. Abbildung 5.1). Die in Abbildung 5.2 dargestellte Altersstruktur der Probanden ist mit einem Durchschnittsalter von 32,8 Jahren bei einer (deskriptiven) Standardabweichung von 2,7 recht homogen. Alle Testteilnehmer sind deutsch und haben (trotz der Durchführung eines Teils der Experimente in Muscat) eine Sozialisierung mit westlicher Musikkultur erfahren.

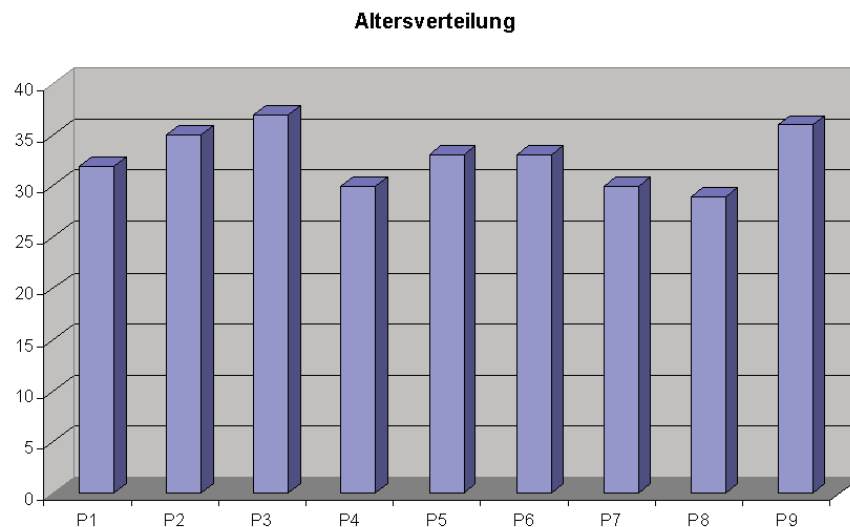


Abbildung 5.2: Alters-Verteilung in der Versuchsgruppe

Neben diesen persönlichen Daten wurden in einem ersten Fragebogen die Hörgewohnheiten und das Musik-Kaufverhalten der Teilnehmer erfasst (vergl. Anhang B).

Bezüglich der bevorzugten Musik-Genres weist die Gruppe ein große Heterogenität auf. Einige Testpersonen bevorzugten Jazz-, HipHop- und Rock-Musik, andere dagegen präferieren eher klassische Musik. Bemerkenswert ist die Vielfältigkeit der Genre-Nennungen innerhalb einer Antwort. So gab Probandin 4 an, regelmäßig Musik der Genres Jazz, Rock und Klassik zu konsumieren. Dies unterstützt die in Abschnitt 4.2.1 getroffene Aussage, dass sich persönlicher Musikgeschmack ohne weiteren Kontext nicht durch Modelle mit singulärem Interessensschwerpunkt darstellen lässt und motiviert den Aufbau einer Explorationsumgebung, die die Musikkollektion des Anwenders zwar strukturiert und damit zugreifbarer macht, letztlich aber dem Nutzer die Suche nach einem gewünschten Titel überlässt.

Ferner wurde die Anzahl an Stunden erfasst, die der Proband durchschnittlich pro Woche Musik hört. Abbildung 5.3 illustriert die Antworten auf diese Frage und zeigt mit einem Mittelwert von 8 Stunden die Bedeutung von Musikkonsum für den Alltag der Versuchsgruppe. Wie sich beim einem Vergleich mit Abbildung 5.4 erkennen lässt, ist kein direkter Zusammenhang zwischen der Hördauer und der Anzahl CDs im Besitz der Probanden nachweisbar. So besitzen P5 und P9 mit 600 CDs mehr Tonträger als P4, hören aber nur halb so viel Musik. In einem weiteren Fragenkomplex wurde die Vertrautheit der Versuchsteilnehmer mit Internet-Musikportalen und digitaler Musik ermittelt. Abbildung 5.5 zeigt zunächst die Anzahl (physikalischer) Alben, die in einem Online-Shop (wie z.B. Amazon.com) erworben wurden. In Abbildung 5.6 ist die Anzahl der Stücke aufgeführt, die in Download-Portalen gekauft wurden. Schließlich wird in Abbildung 5.7 die Anzahl der MP3s dargestellt, die sich in der Musikkollektion der Probanden befinden.

Die Diagramme zeigen deutlich, dass die Testgruppe zwar mit dem Erwerb von physikalischen Tonträgern im Online-Shop vertraut ist (durchschnittlich wurden knapp 30 CDs online erworben). Das selektive Herunterladen von komprimierten MP3s wird jedoch deutlich weniger genutzt (durchschnittlich 6,4 Stücke); die meisten Versuchsteilnehmer haben noch gar keine Erfahrungen mit Download-Portalen gesammelt.

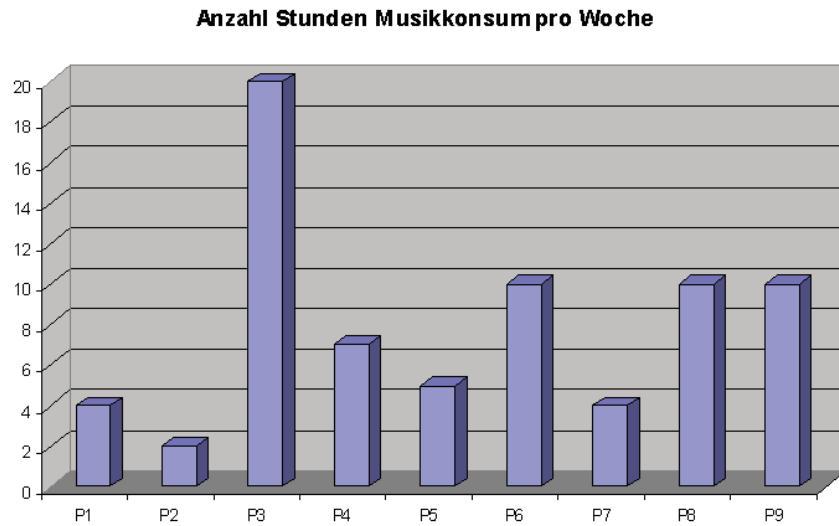


Abbildung 5.3: Anzahl an Stunden, die der Proband durchschnittliche pro Woche mit dem Hören von Musik verbringt

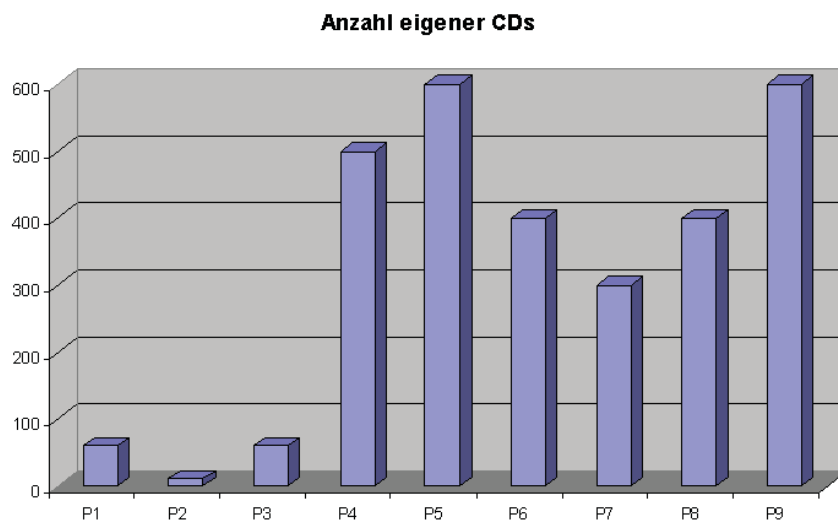


Abbildung 5.4: Anzahl CDs in persönlichem Besitz

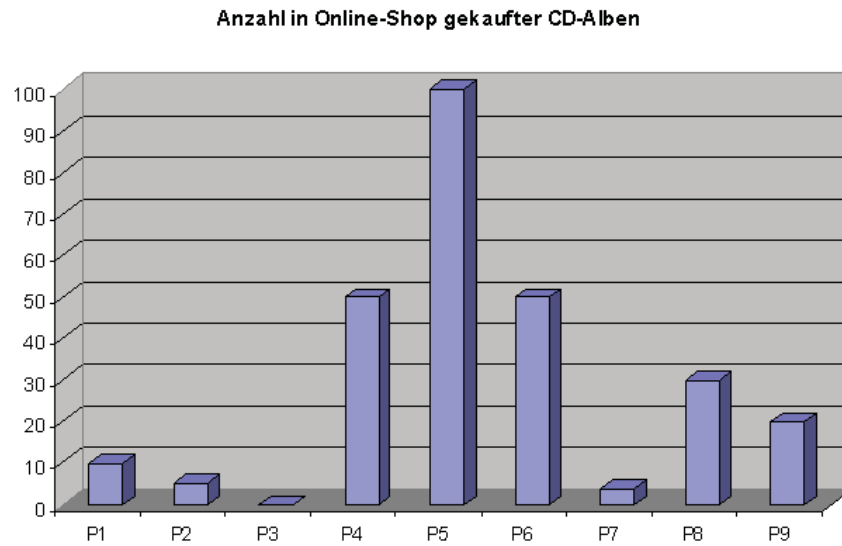


Abbildung 5.5: Anzahl physikalischer Tonträger im Besitz der Probanden

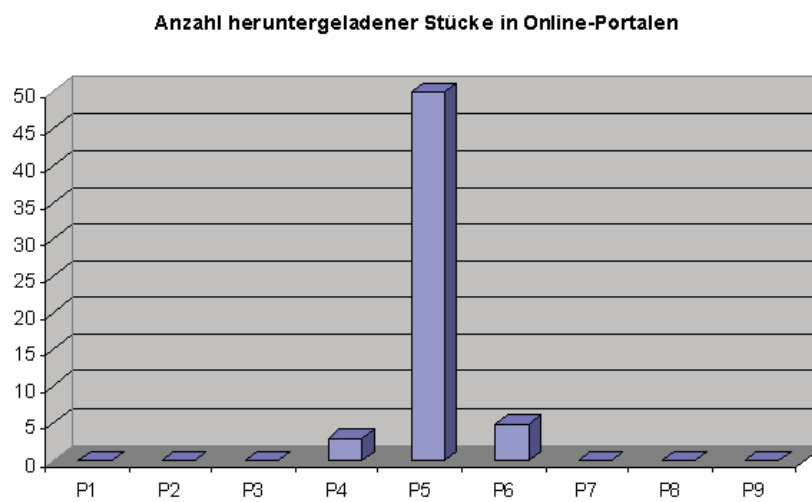


Abbildung 5.6: Anzahl in Online-Shops heruntergeladener Stücke

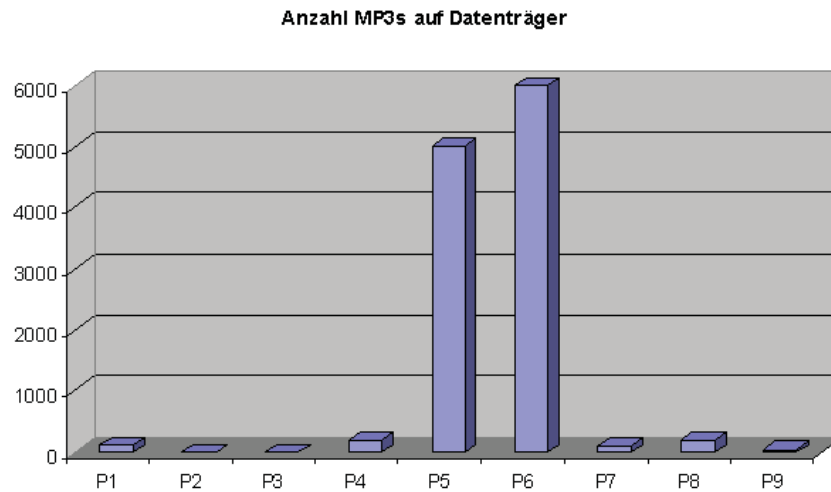


Abbildung 5.7: Größe der MP3-Kollektionen der Probanden

Die geringe Nutzung von MP3s spiegelt sich ebenfalls in der Kollektionsgröße digitaler Musik der Probanden wider. Neben den herausstechenden Werten für P5 und P6 bewegt sich die angegebene Anzahl von MP3s im Besitz der Testpersonen im Bereich von 0 bis 200.

Paarweise Korrelationstests zeigen bei einem Signifikanzniveau von 5% einen linearen Zusammenhang zwischen

- der Anzahl CDs und der Anzahl online erworbener Tonträger (p-Wert: 2,96 %),
- der Anzahl online erworbener Alben und der Anzahl heruntergeladener Stücke (p-Wert: 0,29 %) sowie
- der Anzahl online erworbener Alben und der Größe der privaten MP3-Kollektion (p-Wert: 1,85 %).

Bemerkenswert ist u.a. das Fehlen einer nachweisbaren Abhängigkeit zwischen der Anzahl an CDs im Besitz des Probanden und der Größe seiner MP3-Sammlung.

Aufgrund der recht geringen Anzahl an Evaluationsteilnehmern und der Homogenität der Altersstruktur sind diese Zusammenhänge jedoch nur begrenzt verallgemeinerbar.

Insgesamt lässt sich konstatieren, dass die Versuchsgruppe eher geringe Erfahrungen im Umgang mit digitaler Musik besitzt. Eine Erweiterung der Evaluation um Teilnehmer einer anderen Altersstruktur würde dieses Bild möglicherweise verändern.

Die visuelle und akustische Darstellung einer Musikkollektion im *soniXplorer* erinnert bewusst an aktuelle Computerspiele. Um die Aussagen der Testpersonen zur Benutzerfreundlichkeit des Systems besser beurteilen zu können, wurde daher die Vertrautheit der Probanden mit Computerspielen erfragt. Hierzu mussten die Versuchsteilnehmer den Grad ihrer bisherigen Erfahrung mit Computerspielen auf einer Skala von 1 (keinerlei Erfahrung) bis 7 (sehr erfahren) bewerten. Abbildung 5.8 zeigt ein recht heterogenes Bild der Gruppe. Mit einem Median von 2 ist die durchschnittliche Vertrautheit mit Computerspielen jedoch eher gering.

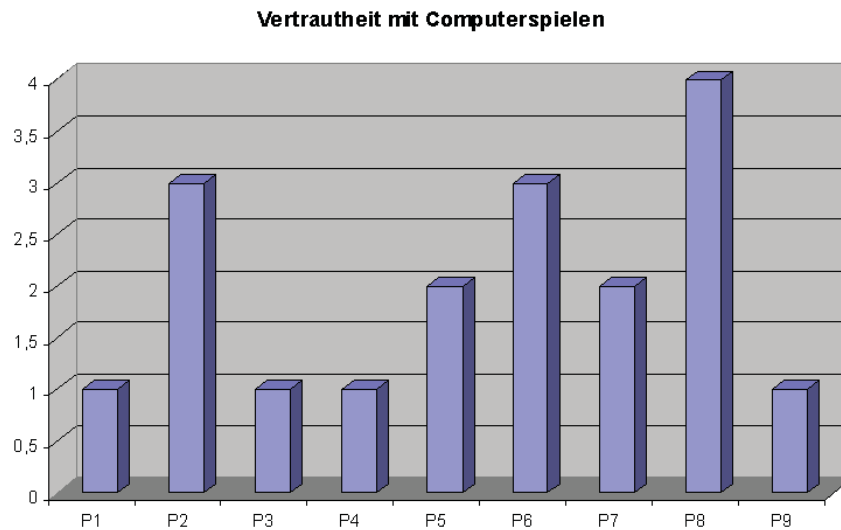


Abbildung 5.8: Vertrautheit der Probanden mit Computerspielen

5.3.2 Quantitative Auswertung

Wie im Entwurfsabschnitt beschrieben, wurde den Probanden zunächst die Gelegenheit gegeben, die Explorationsumgebung nach einer kurzen Einführung in das System selbständig anhand einer unabhängigen Musikkollektion zu erforschen. Nach ca. 15 Minuten fühlten sich alle Versuchsteilnehmer hinreichend vertraut mit der Funktionalität, um mit den eigentlichen Evaluationsexperimenten zu beginnen.

Orientierungsgewinn durch SOM-Strukturierung

Wie erwartet konnte bei allen Probanden eine Verkürzung der Suchzeit nach akustisch präsentierten Stücken in Umgebungen gemessen werden, in denen die Titel nach Ähnlichkeit gruppiert und platziert sind. Abbildung 5.9 stellt die relative Verkürzung dieser Suchzeit für alle neun Versuchsteilnehmer dar.

Wie sich erkennen lässt, ist der Effekt der SOM-Strukturierung zum Teil eher moderat (Verringerung der Suchzeit um 16% bei Proband P3), zum Teil recht groß (78% bei Proband P9). Im Durchschnitt führt die strukturierte Anordnung der Stücke zu einer Verkürzung der Suchzeit um 59% gegenüber der benötigten Zeit bei einer zufälligen Positionierung.

Die Darstellung der *relativen* Suchzeit-Verkürzung trägt dem Umstand Rechnung, dass die Versuchgruppe eine recht große Varianz in der absoluten Dauer der Aufgabenerfüllung zeigt: Während P3 für beide Varianten durchschnittlich 405 Sekunden benötigte, hatte P7 die gesuchten Titel erst nach durchschnittlich 102 Sekunden lokalisiert.

Der Versuch, diese Unterschiede in Bezug zu charakteristischen Merkmalen der Personen (wie z.B. ihre Erfahrung mit Computer-Spielen) zu setzen, ergab keine signifikanten Korrelationen.

Die Ergebnisse dieses Experimentes lassen eine Bestätigung der ersten Evaluations-Hypothese zu, die die Relevanz einer strukturierten Positionierung von Stücken für die Orientierung des Benutzers in der Explorationslandschaft betont.

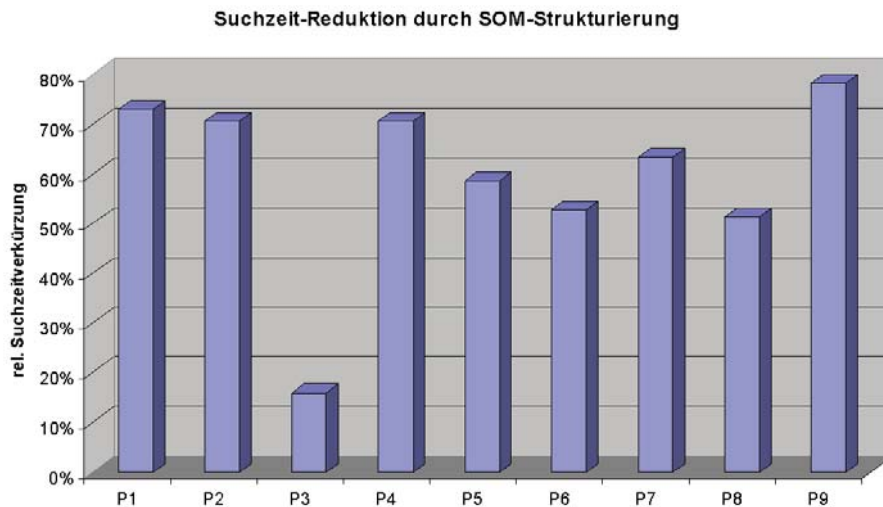


Abbildung 5.9: Reduktion der Suchzeit nach präsentiertem Stück bei verfügbarer SOM-Strukturierung gegenüber zufälliger Verteilung auf Karte

Ferner bestätigen sie die Resultate einer im Kontext dieser Arbeit durchgeführten Studie von Neis, die eine ähnliche Fragestellung für einen Vorläufer-Prototypen des soniXplorers untersucht [Nei07]. Die dort berichtete Suchzeitverkürzung ist zwar klar erkennbar, jedoch deutlich geringer als oben beschrieben.

Orientierungsgewinn durch akustischen Explorationsraum

Der zweite Versuch untersucht die Relevanz der akustischen Wiedergabe für die Orientierung des Benutzers in der Musiklandschaft. Abbildung 5.10 illustriert die gemessene relative Verringerung der benötigten Zeit zur Lokalisierung eines vorgegebenen Stückes durch permanente Auralisation. Wie sich hieraus erkennen lässt, führt die multimodale Darstellung des soniXplorers wie erwartet zu einer Reduktion der Suchzeit um durchschnittlich 58%.

Der erkennbare Orientierungsgewinn durch die zusätzliche Bereitstellung akustischer Informationen steht in Übereinstimmung zu den Resultaten eines Experimentes von Neis [Nei07]: Dort wird jedoch lediglich festgestellt, dass eine rein akustische Exploration ohne visuelle Hilfestellungen machbar und bedienbar ist; eine Untersuchung des Effektes der Ergänzung einer Visualisierung durch akustischer Wiedergabe auf die Orientierungsfähigkeit des Anwenders findet nicht statt.

Anpassungsfähigkeit an unterschiedliche Ähnlichkeitsverständnisse

Im letzten quantitativen Versuch wurden die Probanden zur Umstrukturierung der Kollektion gemäß ihrem subjektiven Ähnlichkeitsempfinden aufgefordert und die Anpassungsfähigkeit des soniXplorers an dieses personalisierte Ordnungsprinzip ermittelt.

Abbildung 5.11 zeigt die ermittelten Platzierungsfehler ohne Anpassung $\widehat{e}_u$ und mit Anpassung $\widehat{e}_a$. Die dort dargestellten Werte sind im Vergleich zum erwarteten Fehler zu interpretieren.

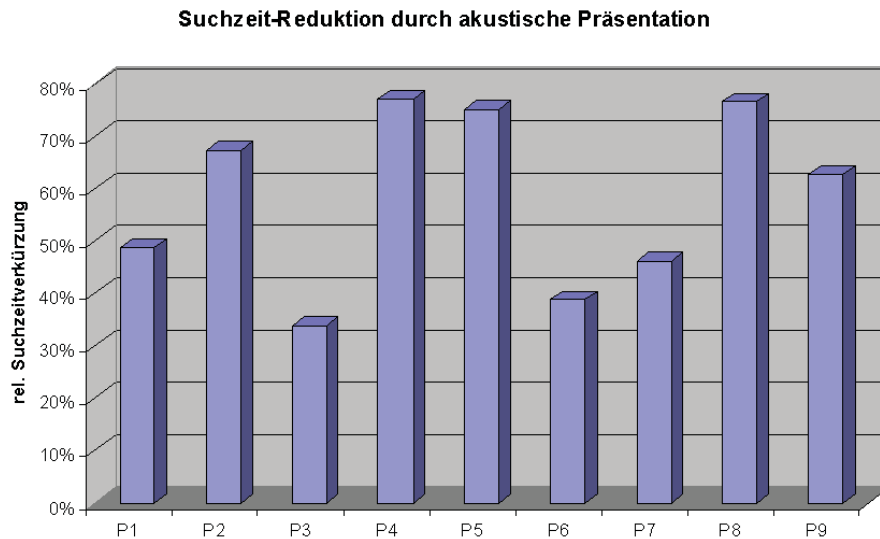


Abbildung 5.10: Reduktion der Suchzeit nach präsentiertem Stück bei permanenter akustischer Umgebung gegenüber benutzergesteuerter Einzel-Wiedergabe

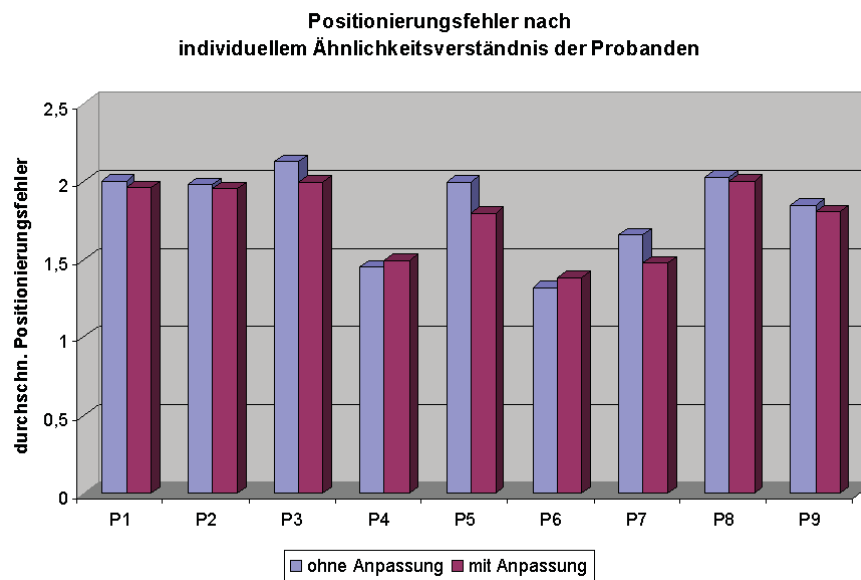


Abbildung 5.11: Durchschnittlicher Positionierungsfehler pro Proband mit und ohne Anpassung an subjektives Ähnlichkeitsverständnis

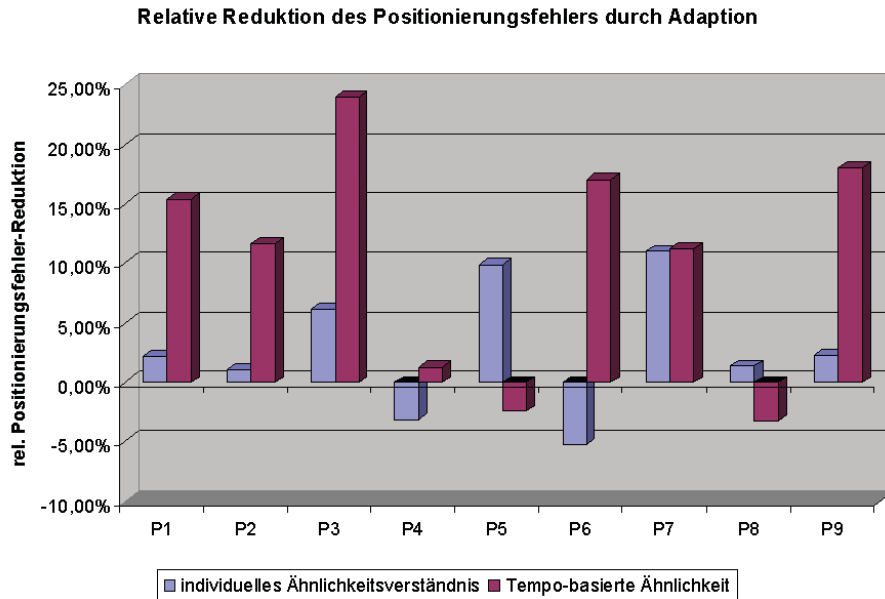


Abbildung 5.12: Relative Reduktion des Positionierungsfehlers durch Adaption bei unterschiedlichen Ähnlichkeitsverständnissen

tieren, der sich aus einer vollständig zufälligen Positionierung der Stücke ergibt:

$$\widehat{e_{random}} = \frac{1}{(a \cdot b)^2} \sum_{\vec{x}} \sum_{\vec{y}} \|\vec{x} - \vec{y}\|,$$

wobei a und b Länge und Breite der Karte darstellen und $\vec{x}$ und $\vec{y}$ Koordinaten dieser Karte sind. Im betrachteten Fall gilt mit $a = b = 20$: $\widehat{e_{random}} \approx 8,06$.

Es zeigt sich, dass die automatische Positionierung ohne Anpassung der Ähnlichkeitsfunktion (d.h. auf der Grundlage des generischen initialen Ähnlichkeitsverständnisses) bereits deutlich besser als eine zufällige Platzierung ist. Gleichzeitig lässt sich jedoch erkennen, dass die Anpassung eine nur geringfügige Reduktion des Platzierungsfehlers bewirkt (siehe Abbildung 5.12). Für die bereits besonders niedrigen Positionierungsfehler bei Proband P4 und P6 führt die Gewichts-Adaption sogar zu einer Verschlechterung der automatischen Stück-Platzierung.

Insgesamt lässt sich aus den Ergebnissen schließen, dass das generische Ähnlichkeitsmaß offenkundig bereits recht gut das subjektive Ähnlichkeitsverständnis der Versuchsteilnehmer repräsentiert und eine Anpassung daher nur marginale Veränderungen bewirkt.

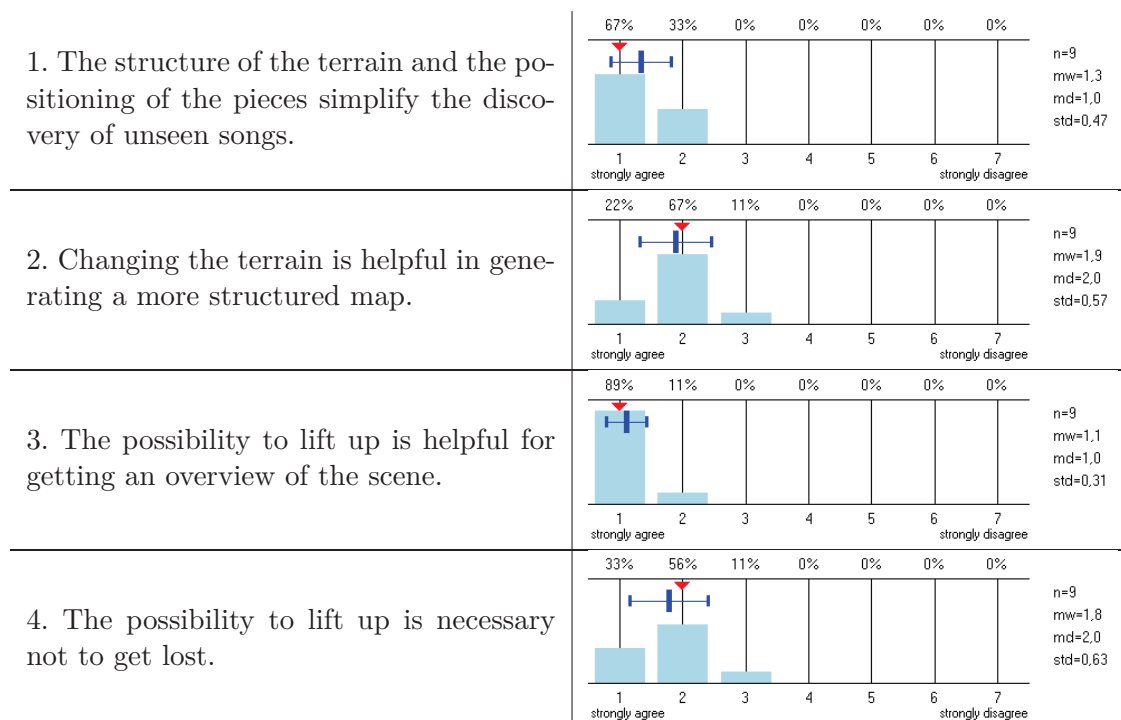
Wie in Abschnitt 5.2.1 beschrieben, wurden die Probanden anschließend gebeten, die Stückkollektion gemäß einem Tempo-dominierten Ähnlichkeitsverständnis umzustrukturieren.

Die Auswertung von analogen Leave-One-Out-Experimenten zeigt nun, dass ohne eine Anpassung der Zusammensetzung des Ähnlichkeitsbegriffes der Positionierungsfehler signifikant höher ist. Abbildung 5.12 fasst dies anschaulich zusammen: Wie sich dort erkennen lässt, führt eine Adaption an das Tempo-basierte Ähnlichkeitsmaß in fast allen Fällen zu einer zum Teil deutlichen Reduktion des relativen Positionierungsfehlers $\frac{e_u - e_a}{e_u}$.

5.3.3 Subjektive Bewertung

In den folgenden Abschnitten werden die Ergebnisse der Fragebogen-Auswertung in den einzelnen Themenfeldern detailliert vorgestellt. Hierzu werden pro Aussage neben einem Histogramm der der geäußerten Zustimmungsniveaus die Anzahl der insgesamt abgegebenen Meinungsäußerungen (n), Mittelwert (mw), Median (md) und Standardabweichung (std) visualisiert.

5.3.4 Strukturierte Kollektionsdarstellung



Die quantitativ in Abschnitt 5.3.2 nachgewiesene Vereinfachung der Exploration einer Musikkollektion durch eine strukturierte Darstellung der Stücke in einer Landschaft aus Tälern und Hügeln wird durch die Reaktionen der Probanden auf Aussage 1 (mw=1,3) deutlich bestätigt.

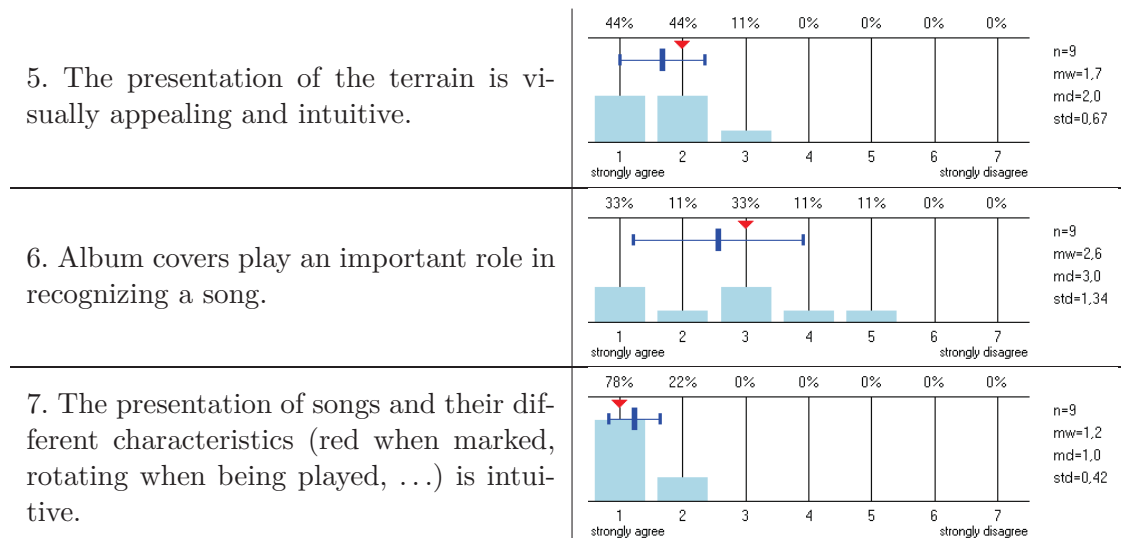
Etwas weniger klar äußern sich die Versuchsteilnehmer zum Effekt einer individuellen Höhenzugs-Veränderung durch den Benutzer. Bei einem Mittelwert von 1,9 stimmt jedoch immer noch eine deutliche Mehrzahl der Probanden der Aussage zu, dass die Möglichkeit zur Anpassung der Landschaft bei der weitergehenden Strukturierung der Karte hilfreich ist.

Ein zentrales Ergebnis der Nutzerstudie, die im Anschluss an die Entwicklung eines *soniXplorer*-Vorgängers durchgeführt wurde, war der Wunsch der Anwender nach einer ergänzenden überblicksartigen Darstellung der Landschaft [Nei07].

Dieser Anforderung wurde im *soniXplorer* durch die Möglichkeit, die Betrachterposition stufenlos vom Grund zu lösen und die Karte so aus einer Art Vogelperspektive zu überblicken,

Rechnung getragen. Die Reaktionen der befragten Benutzer auf die Aussagen 3 und 4 be-
scheinigen die Effektivität dieser Implementierung und bestätigen (wenn auch im Vergleich
zur ersten Benutzerstudie in abgeschwächter Stärke) erneut die Bedeutung dieser Funktionalität.

5.3.5 Visuelle Darstellung

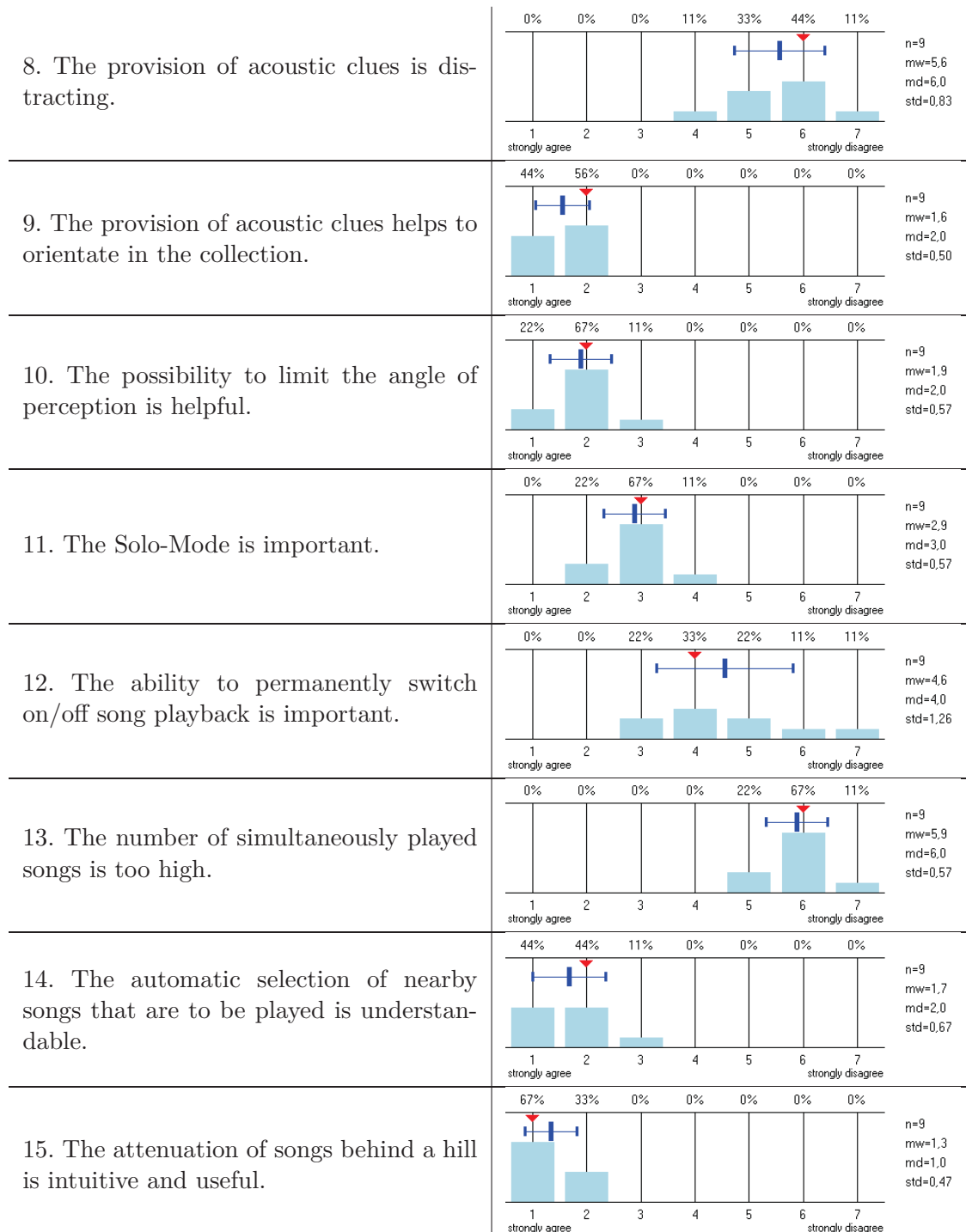


In der nächsten Themengruppe wurde die Meinung der Versuchsgruppe zur visuellen Darstellung der Musikkollektion erfasst. Insgesamt zeigt sich eine recht große Zufriedenheit mit der optischen Präsentation, insbesondere mit der visuellen Kennzeichnung verschiedener Modi der Stücke. In einem Freitextfeld wurde insbesondere die intuitive Verständlichkeit der teilweisen Versenkung von Titeln im Boden, die explizit von der Wiedergabe ausgeschlossen werden, positiv hervorgehoben.

Wie aus den Reaktionen auf Aussage 6 ersichtlich, ordnen die Teilnehmer der Versuchsgruppe der Texturierung von Stück-Würfeln mit Alben-Covern eine unterschiedliche Bedeutung zu. Dies steht in Übereinstimmung mit den Beobachtungen der Probanden, die während der Erfüllung der Suchaufgaben zur quantitativen Evaluation gemacht werden konnten: Während einige Testpersonen sich anhand der wiedergegebenen Stücke rein akustisch in der Umgebung orientierten, versuchten andere, aus der schneller zugreifbaren optischen Präsentation der Alben-Cover eine akustische Vorstellung von Kartenregionen abzuleiten und dies zur gezielten Navigation in der Landschaft zu nutzen.

Dies unterstreicht die Bedeutung der Kombination akustischer und visueller Information in einer Explorationslandschaft und lässt den multimodalen Ansatz des *soniXplorers* als einen zentralen Fortschritt gegenüber bisherigen Systemen erscheinen.

5.3.6 Akustische Darstellung



Die Integration akustischer Orientierungshilfen in die Explorationslandschaft wurde in einem unabhängigen Themenkomplex genauer evaluiert.

Generell beurteilen die Befragten die angebotene akustische Information als wenig verwirrend (Aussage 8) und hilfreich zur Orientierung in der Landschaft (Aussage 9).

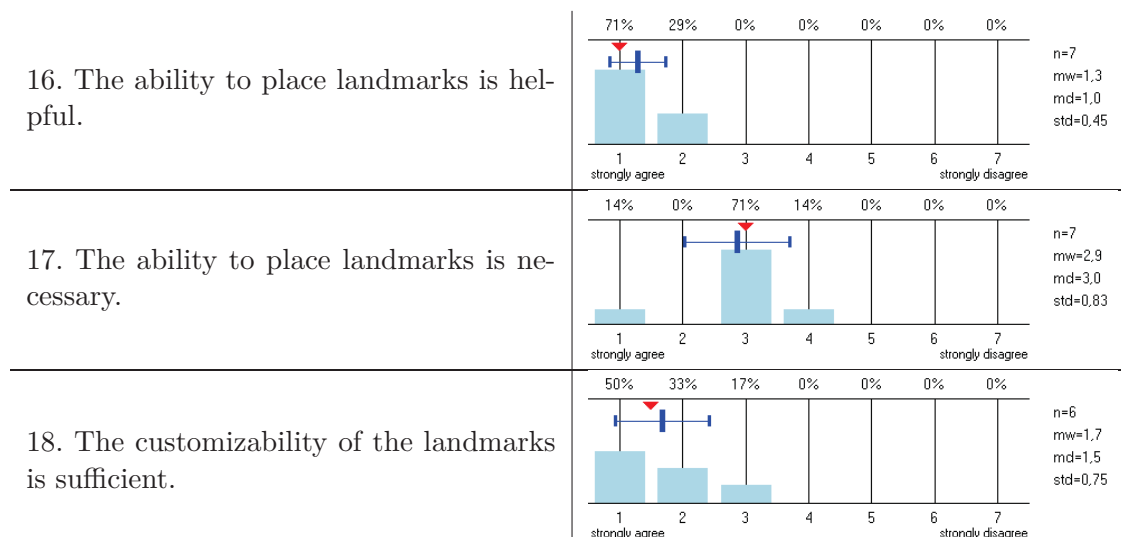
Während des Umgangs mit dem **soniXplorer** machten die Probanden intensiven Gebrauch von der Möglichkeit der akustischen Fokussierung auf das frontale Wahrnehmungsfeld. Diese Einschränkung des „Angle of Perception“ wird von den meisten Teilnehmern positiv bewertet (Aussage 10).

Die Möglichkeit zur vollständigen Ausblendung anderer Stücke und isolierten Wiedergabe eines einzelnen Titels (Solo-Modus) wurden von den Probanden als weniger relevant eingestuft und während der Suchaufgaben kaum genutzt (Aussage 11). Als noch irrelevanter beurteilen die Versuchsteilnehmer die Funktionalität des **soniXplorers** zum expliziten Eingriff in die automatische Auswahl gleichzeitig erklingender Titel (Aussage 12). Dieser Eindruck deckt sich mit der offenkundigen Zufriedenheit der Benutzer mit dem Verfahren, die sich in den Reaktionen auf die Aussagen 13 und 14 zeigt.

Als besonders intuitiv und hilfreich wird von den Probanden die akustische Abschattung von Klangquellen durch verdeckende Höhenzüge wahrgenommen (Aussage 15).

Mehrere Benutzer äußerten im Freitextfeld den Wunsch, die Wiedergabe von Stücke auf „charakteristische Passagen“ zu beschränken, die eine schnelle Beurteilung des Titels „im Vorübergehen“ ermöglichen würde. Ferner beklagte eine Probandin deutliche Lautheits-Unterschiede z.B. zwischen klassischen Stücken und aktuellen Pop-Songs, die eine gleichberechtigte Wahrnehmung der Titel erschweren würde und zu einer vermehrten Nutzung des Fokussierungs-Mechanismus' führten. Offensichtlich ist die in Abschnitt 3.4.2 beschriebene Angleichung der RMS-Werte nicht hinreichend und sollte durch eine Kompression dynamischer oder Expansion weniger dynamischer Stücke ergänzt werden.

5.3.7 Hinweisschilder



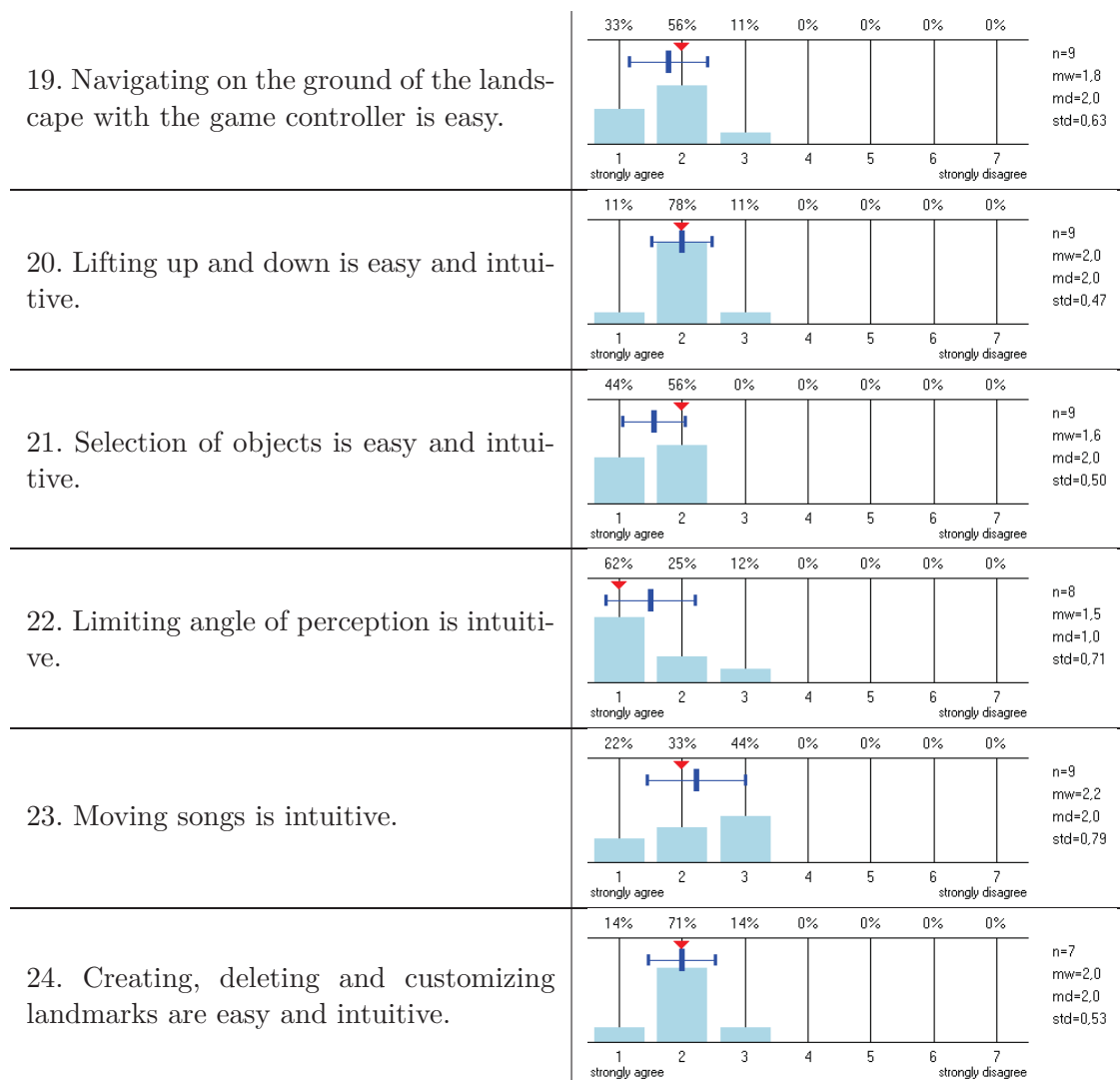
Um die Orientierung des Benutzers in der virtuellen Landschaft zu verbessern, erlaubt der soniXplorer die Kennzeichnung beliebiger Kartenregionen durch die Platzierung frei beschreibbarer Hinweisschilder.

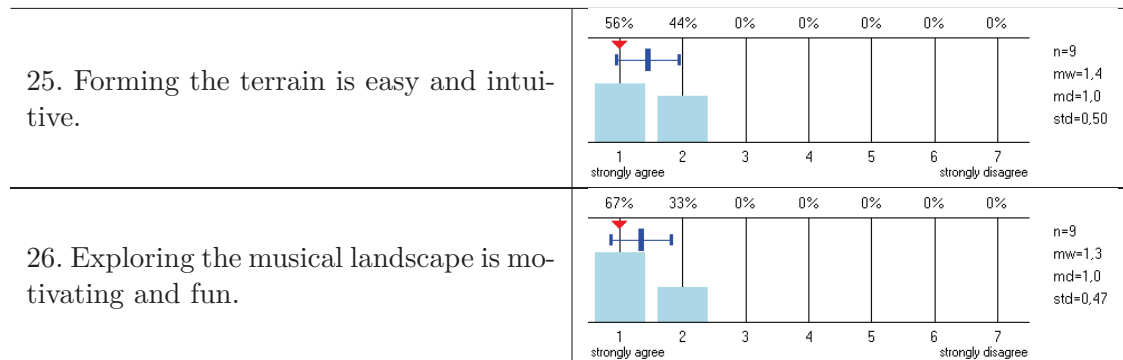
Während die überwiegende Zahl der Versuchsteilnehmer diese Möglichkeit als hilfreich einstuft (Aussage 16), wurde diese Funktionalität von einer kleineren Gruppe als notwendig eingestuft (Aussage 17).

Die meisten Probanden bewerten die verfügbaren Möglichkeiten zur individuellen Anpassung der Hinweisschilder als ausreichend (Aussage 18).

Im freien Kommentar- und Vorschlagsfeld wurde die Idee einer direkt zugreifbaren Bibliothek beschreibender Icons zur Texturierung der Schilder aufgeworfen. Idealerweise solle das System bereits einen Vorschlag zur sinnhaften Annotation der Region unterbreiten.

5.3.8 Allgemeine Benutzerfreundlichkeit





Der letzten Aussagenkomplex schließlich betrifft die wahrgenommene Benutzerfreundlichkeit des Prototypen unter den Probanden. Generell lässt sich eine weitgehende Zufriedenheit der Studienteilnehmer mit der Bedienung des Systems feststellen. Dies gilt für die Navigation in der Umgebung (Aussage 19), die Veränderung der Höhe der Betrachterposition (Aussage 20) sowie für die Selektionsmöglichkeit von Stücken (Aussage 21) und die Steuerung des Wahrnehmungswinkels (Aussage 22). Im Vergleich dazu wird das Verfahren zum Verschieben von Stücken (Selektion des Titels und anschließende Markierung der Zielposition) als weniger intuitiv und benutzerfreundlich eingestuft (Aussage 23).

Im Freitext-Feld wurde auf eine Inkonsistenz zwischen der Bedienung zum Erheben des Beobachters vom Landschaftsgrund und zur Erhöhung bzw. Absenkung des Terrains hingewiesen: Während sich die Betrachterposition mit Hilfe der üblicherweise mit dem Mittelfinger bedienten rechten Trigger-Taste *erhöhen* lässt, führte die Betätigung der rechten Shoulder-Taste (typischerweise mit dem Zeigefinger) zur *Reduzierung* der Terrain-Höhe (vergl. Abbildung 3.26). Diese verwirrende Tastenbelegung konnte durch einen einfachen Tausch der Funktionalität zwischen rechter und linker Shoulder-Taste behoben werden.

Die angebotenen Methoden zum Setzen und Verwalten von Hinweisschildern sowie zur Veränderung des Terrain-Höhenprofils werden durchgehend positiv bewertet (Aussagen 24 und 25).

Eine wichtige Zielsetzung einer Explorationsumgebung ist die Unterstützung des „Entdeckergeistes“ des Benutzers und somit der Förderung seiner Motivation zur Erforschung bislang unbekannter oder wenig gehörter Musik. Die Anlehnung des *soniXplorers* an aktuelle Computerspiele soll diesen Explorationsdrang unterstützen. Wie sich in der deutlichen Zustimmung zu Aussage 26 zeigt, konnte dieses Ziel in einem hohen Maße erreicht werden.

Kapitel 6

Zusammenfassung und Ausblick

Im folgenden werden die zentralen Ergebnisse dieser Arbeit zusammengefasst, die Kern-Resultate herausgestellt und die erreichten Ziele eingeordnet und bewertet. Es schließt sich ein Ausblick auf offene Forschungsfragen und weitere Entwicklungsaufgaben an, die über den gesetzten thematischen Fokus dieser Dissertation hinausreichen oder aufgrund von gesammelten Erfahrungen mit dem entwickelten System sichtbar wurden.

6.1 Zusammenfassung

Wie in Abschnitt 1.2 beschrieben ist das typische Suchverhalten eines Musikinteressenten geprägt durch ein ungerichtetes „Stöbern“ in einer betrachteten Stück- oder Alben-Kollektion. Diesem explorativen Zugriffsmuster steht jedoch zumeist ein auf Funktionen zur gezielten Suche beschränktes Angebot bestehender Musikportale oder Verwaltungssysteme zum Management persönlicher Musiksammlungen gegenüber. Der Bedarf für *Explorationssysteme* wird zudem verstärkt durch die Schwierigkeit, ein musikalisches Informationsbedürfnis adäquat zu artikulieren, wenn das Retrieval-System lediglich assoziierte Metadaten wie Künstlername oder Albumtitel als Zugriffsattribute offeriert.

Diese Beobachtungen motivieren die Konzeption von Systemen zur Unterstützung eines explorativen Zugangs zu Musikkollektionen und bilden die Ausgangsbasis dieser Arbeit. Die Beschreibung einer solchen Konzeptionierung wurde begleitet durch die Diskussion der Entwicklung des Prototypen *soniXplorer*, der eine Evaluation der getroffenen Entwurfsentscheidungen ermöglicht.

Bedienbare Explorationsumgebungen erfordern eine nachvollziehbare Vorstrukturierung der Musiksammlung, die zueinander ähnliche Stücke gruppiert und so eine bessere Orientierung des Anwenders und letztlich eine gezieltere und effizientere Navigation im Explorationsraum ermöglicht.

Daher wurden in Kapitel 2 ausführlich verschiedene Ansätze zur Modellierung eines musikalischen Ähnlichkeitsbegriffes diskutiert und klassifiziert. Zunächst wurde ein vereinheitlichendes Rahmenwerk für *inhaltsbasierte Distanzfunktionen* entwickelt, anhand dessen eine Vielzahl populärer Algorithmen, die unterschiedliche Aspekte wie Melodie, Harmonie oder Rhythmik modellieren, beschrieben und analysiert wurden.

Daran anschließend wurden *Experten-basierte Ähnlichkeitsmaße* vorgestellt, die auf der expliziten Annotation und Eingruppierung durch Musikredakteure oder -wissenschaftler basieren.

Als dritte Gruppe wurden in Abschnitt 2.3 verschiedene zumeist in Online-Portalen genutzte *kollaborative Verfahren* beschrieben, deren Ähnlichkeitsmaße auf der Analyse einer hinreichend großen Anzahl von Nutzeraktivitäten (wie dem Kauf verschiedener Artikel in Internet-Shops, der veröffentlichten Zusammenstellung von Stücken zu Playlists oder der Annotation von Objekten mit frei wählbaren Schlagwörtern) fußen.

Die letzte betrachtete Kategorie bilden *textbasierte Ähnlichkeitsmaße*. Sie reduzieren natürlichsprachliche Beschreibungen (z.B. Rezensionen) der betrachteten Titel, Alben oder Künstler mit Hilfe von Methoden des Natural Language Processing auf informative Beschreibungsvektoren, die ähnlich den beschriebenen Verfahren zur kollaborativen Distanzermittlung ausgewertet werden.

In einer zusammenfassenden Bewertung wurden die Stärken und Schwächen dieser Verfahrensgruppen analysiert (Abschnitt 2.5). So steht der vielfach berichteten hohen Qualität kollaborativer Ähnlichkeitsmaße eine eingeschränkte Anwendbarkeit gegenüber, da eine valide Schätzung eine hinreichend große Datenbasis erfordert. Dagegen sind inhaltsbasierte Distanzen stets berechenbar, erreichen jedoch selten das semantische Niveau eines interpretierenden Hörers („Semantic gap“). Aufgrund dieser konträren Stärken und Schwächen der beiden Ansätze wurde in aufwändigen Experimenten versucht, ein kollaborativen Referenzmaß durch eine Kombination inhaltsbasierter Funktionen zu simulieren (Abschnitt 2.7). Wie dort gezeigt werden konnte, erreicht ein einfaches lineares Modell zwar eine signifikant höhere Korrelation zur kollaborativen Ähnlichkeit als jedes isolierte Einzelmaß; eine vollständig überzeugende Simulation gelingt jedoch nicht.

Zum Aufbau eines umfassenden, Facetten-reichen und gleichzeitig personalisierbaren Ähnlichkeitsmaßes wurde daher eine Reihe von vielfältigen inhaltsbasierten Verfahren im *soniXplorer*-Prototypen linear mit einer Anzahl eigens entwickelter kollaborativer Distanzen kombiniert, die sich aus zugreifbaren Daten des Online-Portals Last.fm und dem Playlist-Publizierer ArtOfTheMix ableiten (Abschnitt 2.8).

Auf der Basis dieses Ähnlichkeitsmaßes wurde im folgenden Kapitel eine Explorationsumgebung konzipiert. Hierzu wird die Matrix paarweiser Abstände zwischen Stücken mit Hilfe Multidimensionaler Skalierung zunächst in einen mittel-dimensionalen Euklidischen Vektorraum transformiert (Abschnitt 3.2). Die so zugeordneten Koordinaten bilden die Eingabe für die Konstruktion einer selbstorganisierenden Merkmalskarte, die die im vorgangehenden Kapitel berechnete Distanz zwischen zwei Titeln durch eine geeignete Positionierung auf der zweidimensionalen Karte approximiert und für den Anwender erfassbarer macht.

Wie durch die in Abschnitt 5.2.1 beschriebene Evaluation bestätigt werden konnte, erleichtert die so gewonnene Strukturierung der Musikkollektion deutlich die Orientierung des Benutzers in der Explorationsumgebung.

Über die Topologie-erhaltende Positionierung von Stücken hinaus ermöglicht die Berechnung der U-Matrix-Werte der SOM-Neuronen die Identifikation von Grenzen, die Kartenbereiche ähnlicher Modellvektor-Zuordnungen trennen. Ein wichtiger Beitrag dieser Arbeit besteht in der Interpretation der U-Matrix-Werte als Höhenangaben einer virtuellen dreidimensionalen Landschaft, die die Platzierung ähnlicher Titel in Tälern und deren Separierung durch Bergketten erlaubt (Abschnitt 3.4.1).

Auf der Basis von Microsofts XNA Framework konnte eine ansprechende Computerspiel-artige Umgebung realisiert werden, in der sich der Benutzer frei bewegen kann. Zur Verstärkung der Immersion des Anwenders mit der virtuellen Landschaft wurde die visuelle Darstellung

durch eine aufwändige akustische HRTF-basierte Präsentation der Stücke in der Umgebung des Benutzers erweitert, die der Gefahr einer akustischen Reizüberflutung aufgrund zu vieler simultan wiedergegebener Titel durch eine benutzergesteuerte Einschränkung des Wahrnehmungswinkels sowie eine intensitätgesteuerte Stück-Selektion begegnet (Abschnitt 3.4.2). Diese Multi-Modalität des Systems vermeidet damit den sonst anzutreffenden Medienbruch, der aus der visuellen (und meist textuellen) Auflistung von akustischen Entitäten (nämlich Musik-Titeln) resultiert. Die Möglichkeit des direkten Hörens erlaubt ein intuitiveres und informierteres Navigieren in der Landschaft. Der positive Effekt der durch diese Arbeit geleisteten zentralen Erweiterung bestehender Explorationsumgebungen auf die Orientierung des Anwenders in der Landschaft konnte in der Abschlussevaluation (Abschnitt 5.2.1) deutlich nachgewiesen werden.

So subjektiv die Wahrnehmung von Musik bei verschiedenen Menschen ist, so unterschiedlich sind ihre nachzunehmenden Ordnungsstrategien, die sich im Kern in verschiedenen Ähnlichkeitsbegriffen äußern. Dieser Heterogenität wurde in dieser Arbeit durch eine adaptive Gewichtung der implementierten Einzeldistanzmaße Rechnung getragen. In Kapitel 4.1 wurden zunächst die realisierten Möglichkeiten zur Anpassung der Explorationsumgebung *durch* den Benutzer beschrieben; hierzu gehören etwa die Platzierung von Hinweisschildern, die Repositionierung von Stücken in der Landschaft sowie die Umgestaltung des Terrain-Höhenprofils. Auf der Grundlage dieser Interaktion kann mit Hilfe eines Multiplen Regressors geschlossen eine Neu-Parametrisierung des linearen Modells ermittelt werden, die die vom Benutzer vorgenommene Umgestaltung der Landschaft bestmöglich erklärt. Die Effektivität dieser Adaption *an* den Benutzer, die einen weiteren Kernbeitrag der Arbeit darstellt, konnte wie in Abschnitt 5.2.1 beschrieben empirisch nachgewiesen werden.

Insgesamt wurde somit ein System zur multimodalen adaptiven Exploration von Musiksammlungen konzipiert, prototypisch in Form des **soniXplorers** implementiert und abschließend evaluiert, das dem Anwender einen intuitiven Zugang zur betrachteten Musikkollektion ermöglicht und ihn durch eine ansprechende visuelle und akustische Darstellung zum Durchstöbern und Entdecken weiterer Titel animiert (vergl. subjektive Bewertung des **soniXplorers** durch die Versuchsteilnehmer in Abschnitt 5.2.2).

6.2 Zukünftige Arbeiten

Die gesammelten Erfahrungen bei Konzeption, Implementierung und Evaluation des beschriebenen Explorationssystems legen eine Reihe viel versprechender Erweiterungsideen nahe und werfen eine Anzahl offener Fragestellungen auf, die im Rahmen dieser Arbeit nicht (mehr) behandelt werden konnten und Gegenstand zukünftiger Forschungsaktivitäten bleiben müssen. Im folgenden werden diese Themen kurz angerissen und ein Ausblick auf mögliche Lösungsansätze gegeben.

6.2.1 Skalierbarkeit

Beschränkung der paarweisen Distanzberechnung

Um die berichtete höhere Qualität von Ähnlichkeitsmaßen nutzbar zu machen, die Musikstücke nicht als Koordinaten eines Euklidischen Vektorraums modellieren, sondern paarweise

Distanzen zwischen Titeln quantifizieren, wird als vereinheitlichende Datenstruktur für jedes implementierte Einzelmaß eine Distanzmatrix aufgestellt, die als Eingabe des linearen Kombinator zu einem personalisierten Ähnlichkeitsmaß dient.

Diese Entscheidung erfordert zum einen die für die SOM-Generierung notwendige anschließende Umwandlung in Vektorraum-Koordinaten durch Multidimensionale Skalierung. Gewichtiger ist jedoch die Tatsache, dass für eine paarweise Distanz-Ermittlung Zeit- und Platzkomplexität der Ähnlichkeitsermittlung quadratisch sind. Dies erschwert eine Skalierung des vorgestellten Ansatzes auf größere Stückmengen. Inwiefern die Verwendung paarweiser Ursprungsdistancen tatsächlich einen Qualitätsgewinn darstellt, der dieses Vorgehen rechtfertigt, konnte im Rahmen dieser Arbeit nicht mehr evaluiert werden.

Eine Möglichkeit, das quadratische Wachstum der Anzahl zu berechnender und zu speichern der Distanzen trotz Beibehaltung der integrierten Ähnlichkeitsmaße einzudämmen, besteht in der Reduktion auf die Betrachtung des Abstandes zu einer geringen Anzahl gewichtiger *Anker-Stücke*, die als repräsentative Prototypen einer homogenen Klasse von ähnlichen Titeln eine näherungsweise Einordnung in Relation zu diesen Referenzpunkten erlauben würden.

Eine ähnliche Idee wird auch von Berenzweig et al. zum Aufbau eines Ähnlichkeitsbegriffes verfolgt, der die Charakterisierung von Stücken (z.B. durch Klang-beschreibende Merkmale) durch ihre Distanz zu festgelegten Referenz-Stücken ersetzt und so eine absolute Merkmalsbeschreibung durch die Berechnung relativer Nähe zu diesen Anker-Titeln substituiert [BEL03].

Es bleibt zu überlegen, welche Stücke der betrachteten Kollektion als Anker-Titel fungieren könnten. Zur Identifikation von Prototypen ließen sich Standard-Clustering-Verfahren anwenden (wie z.B. bei der hierarchischen Gruppierung im SonicRadar (Abschnitt 3.4.2)). Möglicherweise lassen sich für diese Auswahl auch Beobachtungen des Benutzerverhaltens in der Umgebung sinnvoll verwenden: So könnte man aus der manuellen Repositionierung eines Titels durch den Anwender eine bewusste Platzierung des verschobenen Stückes ableiten und damit für eine Eignung als Referenzstück plädieren.

Im soniXplorer wurde die Realisierung einer Reduktion der paarweisen Distanzberechnung bereits vorbereitet: Wie in Abschnitt 3.2 beschrieben lässt sich durch geeignete Wahl der Gewichte w_{ij} im Optimierungskriterium der Multidimensionalen Skalierung die Menge berücksichtigter Unähnlichkeiten δ_{ij} eingrenzen.

Hierarchische selbstorganisierende Karten

Mit der aktuellen Implementation des soniXplorers wurden positive Erfahrungen zur Exploration von Musikkollektionen bis zu einer Größe von etwa 500 Stücken gesammelt. Die Visualisierung umfangreicherer Musikbibliotheken führt jedoch aufgrund der notwendigen Vergrößerung des Terrains zu einer Herabsetzung der Höhenverlaufs-Auflösung (vergl. Abschnitt 3.4.1), um die für eine Immersion erforderliche Reaktivität erhalten zu können. Dieses Problem kann in Zukunft durch eine entfernungsabhängige adaptive Detaillierung der Landschaftsszene gelöst werden, wie sie in aktuellen Computerspielen zur Darstellung komplexer Umgebungen Verwendung findet.

Experimente mit großen virtuellen Explorationsumgebungen zeigten jedoch, dass die Orientierung des Benutzers trotz der Möglichkeit des Setzens von Hinweisschildern mit wachsender Ausdehnung des Terrains zunehmend schwieriger wird und die Anzahl intuitiv erfassbarer und interpretierbarer Objektpositionierungen auf einer selbstorganisierenden Karte begrenzt

ist. Als Ausweg bleibt daher lediglich die Reduktion der darzustellenden Titelmenge. Dies lässt sich wiederum durch Anwendung geeigneter Clustering-Verfahren realisieren, die homogene Gruppen ähnlicher Stücke zu einem repräsentativen Prototyp zusammenfassen. Die Aggregation aller Titel eines Albums zu einer gemeinsamen Entität erscheint aufgrund der mitunter anzutreffenden Verschiedenheit der Stücke einer CD nicht empfehlenswert. Die Zusammenfassung von Stücken zu einem Repräsentanten würde einen hierarchischen Zugriff auf die Sammlung ermöglichen: Die durch den Prototypen beschriebenen Titel könnten in einer unabhängigen Landschaft explorierbar gemacht werden.

Erste Ansätze zur Generierung hierarchischer selbstorganisierender Karte sind bereits von Rauber et al. entwickelt worden [RMD02]. Die Nutzung dieser Verfahren und die Folgen für eine bedienbare Repositionierung von Stücken über Hierarchie-Ebenen hinweg sowie deren Auswirkungen für die Anpassung der personalisierten Ähnlichkeitsfunktion bleiben jedoch Thema zukünftiger Forschung.

6.2.2 Erweiterungen der akustischen Titel-Wiedergabe

Die akustische Präsentation der Stücke in der Umgebung der aktuellen Benutzerposition ermöglicht eine informiertere Navigation in der Explorationslandschaft. Voraussetzung hierfür ist jedoch, dass der Anwender in kurzer Zeit einen repräsentativen Eindruck vom erklingenden Titel erhält. In der bisherigen Implementation des *soniXplorers* wird der gesamte Stückverlauf zyklisch wiedergegeben. Bei einer erneuten Aufnahme eines Titels in die Menge wiederzugebender Stücke wird an der Position fortgefahren, an der das Playback zuvor unterbrochen wurde; dies erleichtert ein Wiedererkennen der Schallquelle durch den Benutzer.

Bei einer Reihe von komplex arrangierten Stücken zeigte sich jedoch, dass die Wiedergabe auch uncharakteristischer Abschnitte des Titels (wie z.B. langer Vor- oder Zwischenspiele) die schnelle Gewinnung eines akustischen Eindrucks erschwert. Daher erscheint die Berücksichtigung des Vorschlags eines Probanden während der Evaluation des zweiten Explorations-Prototypen [Nei07] sinnvoll: An die Stelle einer vollständigen Wiedergabe sollte ein wiederholtes Abspielen eines charakterisierenden Ausschnittes treten. Bei populärer Musik ist dies typischerweise der Refrain des Titels. Zur Vorbereitung dieser Funktionalität wurde in der aktuellen *soniXplorer*-Implementierung bereits die Möglichkeit zur eingegrenzten akustischen Wiedergabe vorgesehen; momentan ist dieser Ausschnitt jedoch manuell festzulegen.

Die automatische Detektion von charakterisierenden Stücksegmenten, die als Zusammenfassung eines Titels fungieren, wird in der Music-Information-Retrieval-Community seit einiger Zeit unter der Bezeichnung *Chorus detection* oder allgemeiner *Audio thumbnailing* intensiv erforscht [CF03, CV03, MK07, PBR02, LC00, Got03, PK06]. Die meisten der publizierten Verfahren identifizieren hierzu wiederkehrende Abschnitte durch eine Selbstähnlichkeitsanalyse unter Verwendung eines geeigneten Ähnlichkeitsbegriffes, der robust gegenüber kleineren Variationen wie leicht veränderten Refrain-Wiederholungen ist. Eine Integration dieser Techniken zur automatischen Identifikation einer kompakten akustischen Repräsentation eines Titels erscheint als viel versprechende Erweiterung der Auralisierung in der Explorationsumgebung.

Während der Evaluation des *soniXplorers* konnte beobachtet werden, dass bei simultaner Wiedergabe benachbarter Stücke ein Titel aufgrund deutlich höherer wahrgenommener Lautheit den anderen akustisch verdeckte. Dieses Problem trat insbesondere dann zu Tage, wenn klassische Stücke zusammen mit aktuellen Pop-Produktionen wiedergegeben wurden. Dies zeigt,

dass die in Abschnitt 3.4.2 diskutierte Angleichung des über den gesamten Stückverlauf gemittelten Lautheitsniveaus aufgrund der großen Unterschiede in der Dynamik verschiedener Titel nicht ausreicht. Daher sollte untersucht werden, inwiefern die Anwendung eines Kompressor-Audio-Effektes [Zö02] zu einer ausgeglicheneren und gleichberechtigteren Wiedergabe führt.

6.2.3 (Semi-)Automatische Annotation von Kartenregionen

Die Kennzeichnung von Regionen der Explorationslandschaft durch benutzerplatzierte Hinweisschilder ermöglicht ein schnelles Wiederfinden markierter Punkte in der Landschaft und erleichtert damit die Orientierung des Anwenders in der Umgebung. Knees et al. nutzen gezielte Web-Suchanfragen, um textuelle Beschreibungen für SOM-Zellen zu generieren, denen eine Stückmenge zugeordnet ist [KPSW07b]. Diese Annotation von Kartenregionen ließe sich gewinnbringend zur (möglicherweise im Dialog mit dem Benutzer abgestimmten) Generierung von Hinweisschildern nutzen.

6.2.4 Alternative Ähnlichkeits-Kombinationsmodelle

Grundlage der Generierung einer selbstorganisierenden Karte ist die Kombination der implementierten Distanzmaße zu einem personalisierten Ähnlichkeitsbegriff. Zur Kombination wird ein einfaches lineares Modell verwendet, das den gewichteten Beitrag der Einzeldistanzen aufsummiert. In zukünftigen Arbeiten sollte untersucht werden, welche Relevanz die einzelnen Distanzen für das effektive Ähnlichkeitsmaß tatsächlich besitzen; aufgrund der Fülle von inhaltsbasierten Merkmalen, die Klangcharakteristika von Stücken betrachten, ist eine hohe Korrelation zwischen verschiedenen dieser Distanzmaßen wahrscheinlich. Die Reduktion redundanter Distanzmaße würde damit die Gefahr einer Überanpassung des linearen Modells an die beschränkte Menge von Trainingsdaten verringern, die sich aus der bewussten Platzierung von Stücken durch den Benutzer ergibt.

Die lineare Kombination ermöglicht eine effiziente Anpassung der Gewichte durch multiple lineare Regression. Andererseits ist die Ausdrucksmächtigkeit der gewichteten Aufsummierung begrenzt. Es bleibt zu untersuchen, ob komplexere Kombinationsmodelle bei einem vertretbaren Adaptionaufwand einen nachweisbaren Mehrwert für die Modellierungsqualität individueller Ähnlichkeitsverständnisse erbringen.

6.2.5 Umfangreiche Evaluationsstudien

Das vorgestellte soniXplorer-System ist der dritte im Kontext dieser Arbeit entwickelte Prototyp zur Musikexploration. In die Konzeption und Implementierung der folgenden Iteration flossen stets Evaluationsergebnisse und Anwender-Anregungen der vorherigen Evolutionsstufe ein. Somit kann man von einer Nutzer-getriebenen Entwicklung sprechen.

Die Aussagekraft der in Kapitel 5 diskutierten Abschlussevaluation könnte jedoch durch eine Vergrößerung der Probandenzahl gesteigert werden. Wie in Abschnitt 5.3.1 beschrieben sind insbesondere die Altersverteilung, das Hörverhalten und die Erfahrung mit Computerspielen der Testpersonen recht homogen. Interessant wäre daher u.a. eine Untersuchung dieser Einflussfaktoren auf die Leistungen in den szenariobasierten Experimenten sowie auf die Zufriedenheit der Anwender mit dem System.

Anhang A

Verwendete Musik-Kollektionen

A.1 Demonstrations-Kollektion

ID	Titel	Künstler
296	Die sein	Annett Louisan
299	Auf dich hab ich gewartet	Annett Louisan
301	So Very Hard to Go	Tower of Power
302	Piano Concerto No. 9	Wolfgang Amadeus Mozart
303	Concerto in E Major	Antonio Vivaldi
318	Unchain My Heart	Joe Cocker
321	Sultans Of Swing	Dire Straits & Mark Knopfler
322	Sing, Sing, Sing	New York Voices
324	Menuett d-Moll	Johann Sebastian Bach
327	Seven Years	Norah Jones
329	Give Me Your Love	Tower of Power
341	Now That the Magic Has Gone	Joe Cocker
350	Oakland Zone	Tower of Power
354	I've Got to See You Again	Norah Jones
357	Don't Let Me Be Misunderstood	Nina Simone
360	Stranger In My Own House	Tower of Power
361	Vielleicht	Annett Louisan
362	Hey!	Die Fantastischen Vier
363	Sinnerman	Nina Simone
366	Troy	Die Fantastischen Vier
367	In A Mellow Tone	New York Voices
372	Mamma Mia	ABBA
374	The Winner Takes It All	ABBA
395	Tease Me Please Me	Scorpions
396	Restless Nights	Scorpions

Fortsetzung auf nächster Seite...

ID	Titel	Künstler
413	Moderato	Pyotr Ilyich Tchaikovsky
414	Theme Et Variations	Pyotr Ilyich Tchaikovsky
420	Lobet den Herrn	Johann Sebastian Bach
429	Spring Symphony (op. 44), Part I	Benjamin Britten
430	Den Tod niemand zwingen kunnt	Johann Sebastian Bach
435	Spring Symphony (op. 44), Part II	Benjamin Britten
441	Spring Symphony	Benjamin Britten
445	Wo soll ich fliehen hin	Johann Sebastian Bach
446	Walk of Life	Dire Straits
450	On Every Street	Dire Straits
451	Sie ist weg	Die Fantastischen Vier
484	Zuhälter	Wir sind Helden
485	Zieh Dir was an	Wir sind Helden
491	Christ Lag In Todesbanden	The Hilliard Ensemble
492	Jesu Deine Passion	The Hilliard Ensemble
500	When I'm Grown	Tok Tok Tok
502	Don't Go Away	Tok Tok Tok
503	I Sing You Sing	The Real Group
508	A Cappella In Acapulco	The Real Group
524	Hab' Erdöl im Garten	Reinhard Mey
525	Wem Gott die rechte Gunst	Reinhard Mey
526	Ich bin Klempner von Beruf	Reinhard Mey
567	African	Various Artists
568	Mi Caan Sleep	Romain Virgo
569	Tonight Is the Night	Nikeisha Barnes
570	Take It Off	Anthony Cruz
571	You Can't Blame the Youth	Bushman

Tabelle A.1: Demonstrations-Kollektion aus Abschnitt 3.4

A.2 Kollektion zur Evaluation der Bedeutung strukturierter Darstellung

ID	Titel	Künstler
310	Kompromisse	Roger Cicero
311	From Holberg's Time - Preludium	Edvard Grieg
312	Just My Imagination	The Cranberries
313	Belfast Child	Simple Minds
314	Nun ruhen alle Wälder	Dieter Falk
315	Like a Star	Corinne Bailey Rae
316	Rockferry	Duffy
317	Stop This World	Diana Krall
318	Unchain My Heart	Joe Cocker
319	Yo George	Tori Amos
320	Besser gehts nicht	2raumwohnung
321	Sultans Of Swing	Dire Straits & Mark Knopfler
322	Sing, Sing, Sing	New York Voices
323	A-Tisket, A-Tasket	Ella Fitzgerald
324	Menuett d-Moll	Johann Sebastian Bach
325	Dreams (Live)	The Cranberries
326	From Holberg's Time - Sarabande	Edvard Grieg
327	Seven Years	Norah Jones
328	Enchantment	Corinne Bailey Rae
329	Give Me Your Love	Tower of Power
330	Klar	Jan Delay
331	Lullaby of Birdland	Ella Fitzgerald
332	It's Going to Take Some Time	Carole King
333	Put Your Records On	Corinne Bailey Rae
334	Temptation	Diana Krall
335	Just Friends	Amy Winehouse
336	Jetzt geht's ab	Die Fantastischen Vier
337	Twentysomething	Jamie Cullum
338	Kirchturmkandidaten	Jan Delay
339	Des Abends	Robert Schumann
340	I've Changed My Address	Diana Krall
341	Now That the Magic Has Gone	Joe Cocker
342	Pipis und Popos	Die Fantastischen Vier
343	Lobet, den Herren	Dieter Falk
344	I'm Pulling Through	Diana Krall
345	Mercy	Duffy
346	Millionen Legionen	Die Fantastischen Vier
347	Wo ist das Problem?	Annett Louisan

Fortsetzung auf nächster Seite...

ID	Titel	Künstler
348	Blame It on My Youth	Jamie Cullum
349	Ich Idiot ließ dich gehen	Roger Cicero
350	Oakland Zone	Tower of Power
351	Wake Up Alone	Amy Winehouse
352	Walk of Life	Dire Straits
353	I'm Scared	Duffy
354	I've Got to See You Again	Norah Jones
355	Serenade, Op.22: Larghetto	Antonín Dvorák
356	Sie ist weg	Die Fantastischen Vier
357	Don't Let Me Be Misunderstood	Nina Simone
358	Devils and Gods	Tori Amos
359	Hypnotised	Simple Minds
360	Stranger In My Own House	Tower of Power
361	Vielleicht	Annett Louisan
362	Hey!	Die Fantastischen Vier
363	Sinnerman	Nina Simone
364	Wrap Around Joy	Carole King
365	Nightingale	Carole King
366	Troy	Die Fantastischen Vier
367	In A Mellow Tone	New York Voices
368	You've Got A Friend	Carole King
369	A Kind of Magic	Queen
370	Posse Bonus	Tori Amos
371	Headlong	Queen
372	Mamma Mia	ABBA
373	Lay All Your Love on Me	ABBA
374	The Winner Takes It All	ABBA
375	Chiquitita	ABBA
376	Teach Me Tonight	Al Jarreau
377	Gimme What You Got	Al Jarreau
378	Marry You	B.B. King & Eric Clapton
379	English Suite No. 2 in A minor	Johann Sebastian Bach
380	English Suite No. 3 in G minor	Johann Sebastian Bach
381	Satin Soul	Love Unlimited Orchestra
382	You Sexy Thing	Hot Chocolate
383	Love's Theme	Love Unlimited Orchestra
384	Can't Get Enough of Your Love, Babe	Barry White
385	You're the First, the Last, My Everything	Barry White
387	Movin' Out (Anthony's Song)	Billy Joel
388	She's Always a Woman	Billy Joel
389	Say Goodbye to Hollywood	Billy Joel

Fortsetzung auf nächster Seite...

ID	Titel	Künstler
390	The Stranger	Billy Joel
391	Before You Accuse Me	Eric Clapton
392	Layla	Eric Clapton
393	Cry Me a River	Ella Fitzgerald
394	Feeling Good	Nina Simone
395	Tease Me Please Me	Scorpions
396	Restless Nights	Scorpions
397	Send Me an Angel	Scorpions
398	Scarborough Fair / Canticle	Simon & Garfunkel
399	Bridge Over Troubled Water	Simon & Garfunkel
400	See the Light	Simple Minds
401	The Logical Song	Supertramp
402	Take the Long Way Home	Supertramp
403	Just Another Nervous Wreck	Supertramp
404	Breakfast in America	Supertramp
405	Dreamer	Supertramp
406	How Deep Is Your Love	Take That
407	Symphony No. 5 (1. Satz)	Gustav Mahler
408	Symphony No. 5 (3. Satz)	Gustav Mahler
409	Symphony No. 5 (4. Satz)	Gustav Mahler
410	Promises	The Cranberries
411	Copycat	The Cranberries
412	Piano Concerto No. 1	Pyotr Ilyich Tchaikovsky
413	Moderato	Pyotr Ilyich Tchaikovsky
414	Theme Et Variations	Pyotr Ilyich Tchaikovsky
415	The Big Picture	Y Kant Tori Read
416	Pirates	Y Kant Tori Read
417	You Go to My Head	Y Kant Tori Read
490	One For My Baby	Ida Sand
491	Christ Lag In Todesbanden	The Hilliard Ensemble
492	Jesu Deine Passion	The Hilliard Ensemble

Tabelle A.2: Kollektion zur Evaluation der Bedeutung strukturierter Kollektions-Darstellung (Abschnitt 5.2.1)

A.3 Kollektionen zur Evaluation der Bedeutung akustischer Wiedergabe

ID	Titel	Künstler
301	So Very Hard to Go	Tower of Power
308	Ohne dich	Wise Guys
309	Root Beer Rag	Wise Guys
322	Sing, Sing, Sing	New York Voices
329	Give Me Your Love	Tower of Power
350	Oakland Zone	Tower of Power
360	Stranger In My Own House	Tower of Power
367	In A Mellow Tone	New York Voices
487	Sing mal wieder	Wise Guys
494	Chili Con Carne (New Version)	The Real Group
503	I Sing You Sing	The Real Group
508	A Cappella In Acapulco	The Real Group
514	Dead End Street	Lou Rawls
515	Twenty Flight Rock	Dick Brave & The Backbeats
516	Hallelujah, I Love Her So	Dick Brave & The Backbeats
517	Ex oriente	Unbekannt
518	Lady Love	Lou Rawls
519	Mittelalter	Unbekannt
535	Don't Speak	Sarah Menescal
536	Myriama	Douzi
537	Run Away (Original Mix)	Dennis Sheperd
538	The End of the Line	Metallica
539	Domination	Symphony X
540	Drift Away (feat. Aneym)	Ronski Speed
541	The Girl For Me Tonight	Peter Cincotti
542	You're Looking At Country	Loretta Lynn
543	Walking After Midnight	Patsy Cline
544	Panic Attack	Dream Theater
545	Less is More	Joss Stone
546	A Boy Named Sue	Johnny Cash
547	Creeps (Ronski Speed Remix)	Sebastian Sand
548	Auf der Mauer, auf der Lauer...	Petra Hemker, Olaf Hemker
549	Every Breath You Take	Karen Souza
550	Hip Shaker	Oojami
551	Blanket On the Ground	Billie Jo Spears
552	Camelride	Cosmic Orient
553	In den Bergen ist's am besten	Peter Wackel
554	Es steht in deinen Augen	Nic

Fortsetzung auf nächster Seite...

ID	Titel	Künstler
555	Ich liebe das Leben	Rabaue
556	Der Mai, der Mai, der lustige	Rolf Zuckowski
557	Die Maus auf Weltraumreise	Volker Rosin
558	Rudolph mit der roten Nase	Rudolph mit der roten Nase
559	River man	Till Brönner
560	Evapor8 (Album Version)	Ronski Speed
561	Paint It, Black	Sixth Finger
562	One Love	Anakelly
563	Der schmucke Prinz	Fritz Weber
564	Heidewitzka Herr Kapitän	Various Artists
565	Mer losse dr Dom en Kölle	Bläck Fööss
566	Superjeilezick	Brings

Tabelle A.3: Erste Kollektion zur Evaluation der Bedeutung permanenter akustischer Wiedergabe (Abschnitt 5.2.1)

ID	Titel	Künstler
296	Die sein	Annett Louisan
299	Auf dich hab ich gewartet	Annett Louisan
301	So Very Hard to Go	Tower of Power
302	Piano Concerto No. 9	Wolfgang Amadeus Mozart
303	Concerto in E Major	Antonio Vivaldi
318	Unchain My Heart	Joe Cocker
321	Sultans Of Swing	Dire Straits & Mark Knopfler
322	Sing, Sing, Sing	New York Voices
324	Menuett d-Moll	Johann Sebastian Bach
327	Seven Years	Norah Jones
329	Give Me Your Love	Tower of Power
341	Now That the Magic Has Gone	Joe Cocker
350	Oakland Zone	Tower of Power
354	I've Got to See You Again	Norah Jones
357	Don't Let Me Be Misunderstood	Nina Simone
360	Stranger In My Own House	Tower of Power
361	Vielleicht	Annett Louisan
362	Hey!	Die Fantastischen Vier
363	Sinnerman	Nina Simone
366	Troy	Die Fantastischen Vier
367	In A Mellow Tone	New York Voices
372	Mamma Mia	ABBA
374	The Winner Takes It All	ABBA
395	Tease Me Please Me	Scorpions

Fortsetzung auf nächster Seite...

ID	Titel	Künstler
396	Restless Nights	Scorpions
413	Moderato	Pyotr Ilyich Tchaikovsky
414	Theme Et Variations	Pyotr Ilyich Tchaikovsky
420	Lobet den Herrn	Johann Sebastian Bach
429	Spring Symphony (op. 44), Part I	Benjamin Britten
430	Den Tod niemand zwingen kunnt	Johann Sebastian Bach
435	Spring Symphony (op. 44), Part II	Benjamin Britten
441	Spring Symphony	Benjamin Britten
445	Wo soll ich fliehen hin	Johann Sebastian Bach
446	Walk of Life	Dire Straits
450	On Every Street	Dire Straits
451	Sie ist weg	Die Fantastischen Vier
484	Zuhälter	Wir sind Helden
485	Zieh Dir was an	Wir sind Helden
491	Christ Lag In Todesbanden	The Hilliard Ensemble
492	Jesu Deine Passion	The Hilliard Ensemble
500	When I'm Grown	Tok Tok Tok
502	Don't Go Away	Tok Tok Tok
503	I Sing You Sing	The Real Group
508	A Cappella In Acapulco	The Real Group
524	Hab' Erdöl im Garten	Reinhard Mey
525	Wem Gott die rechte Gunst	Reinhard Mey
526	Ich bin Klempner von Beruf	Reinhard Mey
567	African	Various Artists
568	Mi Caan Sleep	Romain Virgo
569	Tonight Is the Night	Nikeisha Barnes
570	Take It Off	Anthony Cruz
571	You Can't Blame the Youth	Bushman

Tabelle A.4: Zweite Kollektion zur Evaluation der Bedeutung permanenter akustischer Wiedergabe (Abschnitt 5.2.1)

A.4 Kollektion zur Evaluation der Benutzeradaption

ID	Titel	Künstler
287	Soul With a Capital 'S'	Tower of Power
288	Minuet in D Minor	Johann Sebastian Bach
289	It Had Better Be Tonight	Michael Bublé
290	Torsten Schmidt	Annett Louisan
292	Kein Abendessen	Roger Cicero
295	Für immer und Dich	Jan Delay
296	Die sein	Annett Louisan
297	Goldberg Variations	Johann Sebastian Bach
299	Auf dich hab ich gewartet	Annett Louisan
300	Ich hätt so gern noch Tschüss gesagt	Roger Cicero
301	So Very Hard to Go	Tower of Power
302	Piano Concerto No. 9	Wolfgang Amadeus Mozart
304	After All	Al Jarreau
307	Concerto in A minor	Antonio Vivaldi
309	Root Beer Rag	Wise Guys

Tabelle A.5: Kollektion zur Evaluation der Benutzeradaption (Abschnitt 5.2.1)

Anhang B

Evaluations-Fragebogen

Auf den folgenden Seiten ist der in Abschnitt 5.2.2 zur Evaluation des soniXplorers konzipierte Fragebogen abgedruckt, der neben der Erfassung persönlicher Daten der Probanden deren subjektiven Eindruck vom System erhebt.

Evaluation of soniXplorer – Questionnaire

Date: _____

Personal Info

Name: _____

Sex: ☐ female ☐ male

Age: _____

Music Listening Habits and Purchase Experience

Name the musical genres you're mainly interested in:

How many hours do you spend on listening to music per week (approx.): _____

How many CDs do you own (approx.): _____

How many albums have you bought in online shops so far? _____

How many songs have you downloaded from download portals so far? _____

How many MP3s do you have on your hard disk (approx.): _____

How experienced are you with Computer Games?

☐ ☐ ☐ ☐ ☐ ☐
no experience very experienced

Please indicate for the following statements the degree by which you agree or disagree with them.

<u>Structured Organization of Songs on a Map</u>		
1. The structure of the terrain and the positioning of the pieces simplify the discovery of unseen songs.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
2. Changing the terrain is helpful in generating a more structured map.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
3. The possibility to lift up is helpful for getting an overview of the scene.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
4. The possibility to lift up is necessary not to get lost.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
Comments and suggestions for improvement:		
<u>Visual Presentation</u>		
5. The presentation of the terrain is visually appealing and intuitive.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
6. Album Covers play an important role in recognizing a song.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
7. The presentation of songs and their different characteristics (red when marked, rotating when being played, ...) is intuitive.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
Comments and suggestions for improvement:		

<u>Acoustic Clues</u>		
8. The provision of acoustic clues is distracting.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	I agree I disagree I don't know
9. The provision of acoustic clues helps to orientate in the collection.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	I agree I disagree I don't know
10. The possibility to limit the angle of perception is helpful.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	I agree I disagree I don't know
11. The Solo-Mode is important.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	I agree I disagree I don't know
12. The ability to permanently switch on/off song playback is important.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	I agree I disagree I don't know
13. The number of simultaneously played songs is too high.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	I agree I disagree I don't know
14. The automatic selection of nearby songs that are to be played is understandable.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	I agree I disagree I don't know
15. The attenuation of songs behind a hill is intuitive and useful.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	I agree I disagree I don't know
Comments and suggestions for improvement:		
<u>Landmarks</u>		
16. The ability to place landmarks is helpful.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	I agree I disagree I don't know
17. The ability to place landmarks is necessary.	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	I agree I disagree I don't know
18. The customizability of the landmarks is sufficient	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐	I agree I disagree I don't know
Comments and suggestions for improvement:		

Usability		
19. Navigating on the ground of the landscape with the game controller is easy.	☐ ☐ ☐ ☐ ☐ ☐ ☐ I agree I disagree	☐ I don't know
20. Lifting up and down is easy and intuitive.	☐ ☐ ☐ ☐ ☐ ☐ ☐ I agree I disagree	☐ I don't know
21. Selection of objects is easy and intuitive.	☐ ☐ ☐ ☐ ☐ ☐ ☐ I agree I disagree	☐ I don't know
22. Limiting angle of perception is intuitive.	☐ ☐ ☐ ☐ ☐ ☐ ☐ I agree I disagree	☐ I don't know
23. Moving songs is intuitive.	☐ ☐ ☐ ☐ ☐ ☐ ☐ I agree I disagree	☐ I don't know
24. Creating, deleting and customizing landmarks are easy and intuitive.	☐ ☐ ☐ ☐ ☐ ☐ ☐ I agree I disagree	☐ I don't know
25. Forming the terrain is easy and intuitive.	☐ ☐ ☐ ☐ ☐ ☐ ☐ I agree I disagree	☐ I don't know
26. Exploring the musical landscape is motivating and fun.	☐ ☐ ☐ ☐ ☐ ☐ ☐ I agree I disagree	☐ I don't know
Comments and suggestions for improvement:		
Overall comments and suggestions for improvement:		

Literaturverzeichnis

- [ADR04] Miguel Alonso, Bertrand David, and Gaël Richard. Tempo and Beat Estimation of Musical Signals. In *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR)*, Barcelona, Spain, 2004.
- [AF04] J.-J. Aucouturier and Pachet F. Improving Timbre Similarity: How High Is the Sky? *Journal of Negative Results in Speech and Audio Sciences*, 1(1), 2004.
- [Ahr05] P. Ahrendt. The Multivariate Gaussian Probability Distribution. Technical report, Technical University of Denmark, 2005.
- [ALdM01] J.L. Arcos and R. López de Mántaras. An Interactive CBR Approach for Generating Expressive Music. *Journal of Applied Intelligence*, 14(1):115–129, 2001.
- [AP02] Jean-Julien Aucouturier and Francois Pachet. Finding Songs that Sound the Same. In *Proceedings of IEEE Benelux Workshop on Model-Based Processing and Coding of Audio*, Leuven, Belgium, November 2002.
- [AP03] Jean-Julien Aucouturier and François Pachet. Representing Musical Genre: A State of the Art. *Journal of New Music Research*, 32(1), 2003.
- [Aro92] B. Arons. A Review of the Cocktail Party Effect. *Journal of the Americal Voice I/O Society*, 12:35–50, 1992.
- [Bau03] Stephan Baumann. Music Similarity Analysis in a P2P Environment. In *Proceedings of the 4th European Workshop on Image Analysis for Multimedia Interactive Services*, London, UK, April 2003.
- [Bau05a] Stephan Baumann. *Artificial Listening Systems*. PhD thesis, Technische Universität Kaiserslautern, 2005.
- [Bau05b] Stephan Baumann. Tutorial Visualization for Music IR. In *6th International Society for Music Information Retrieval Conference (ISMIR)*, London, England, 2005.
- [BEL03] A. Berenzweig, D. P. W. Ellis, and S. Lawrence. Anchor Space for Classification and Similarity Measurement of Music. In *ICME '03: Proceedings of the 2003 International Conference on Multimedia and Expo*, pages 29–32, Washington, DC, USA, 2003. IEEE Computer Society.

- [BF03] Eoin Brazil and Mikael Fernström. Audio Information Browsing with the Sonic Browser. In *CMV '03: Proceedings of the conference on Coordinated and Multiple Views In Exploratory Visualization*, Washington, DC, USA, 2003. IEEE Computer Society.
- [BFTC02] Eoin Brazil, Mikael Fernström, George Tzanetakis, and Perry Cook. Enhancing Sonic Browsing Using Audio Information Retrieval. In *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan, 2002.
- [BH03] Stephan Baumann and Oliver Hummel. Using Cultural Metadata for Artist Recommendations. In *Proceedings of the Third International Conference WEB Delivering of Music (WEDELMUSIC 03)*, 2003.
- [Bla04] David Blackburn. Does File Sharing Affect Record Sales? Working Paper, Harvard University, Department of Economics, Cambridge, Mass., April 2004.
- [BLEW04] Adam Berezweig, Beth Logan, Daniel P.W. Ellis, and Brian Whitman. A Large-Scale Evaluation of Acoustic and Subjective Music Similarity Measures. *Computer Music Journal*, 28(2):63 – 76, 2004.
- [BP03] Stephan Baumann and Tim Pohle. A Comparison of Music Similarity Measure for a P2P Application. In *Proceedings of the Sixth International Conference on Digital Audio Effects DAFX*, London, UK, September 2003.
- [BPV04] Stephan Baumann, Tim Pohle, and Shankar Vembu. Towards a Socio-Cultural Compatibility of MIR Systems. In *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR)*, pages 460 – 465, Barcelona, Spain, 2004.
- [Bra99] Karlheinz Brandenburg. MP3 and AAC Explained. In *AES 17th International Conference on High Quality Audio Coding*, 1999.
- [Bri92] E. Brill. A Simple Rule-based Part-of-Speech Tagger. In *Proceedings of ANLP-92, 3rd Conferenc on Applied Natural Language Processing*, pages 152–155, Trento, Italy, 1992.
- [BSJ00] A. Becks, S. Sklorz, and M. Jarke. A Modular Approach for Exploring the Semantic Structure of Technical Document Collections. In *ACM Proc. of the 5th Intern. Working Conf. on Advanced Visual Interfaces (AVI'2000)*, pages 298–301, Palermo, Italy, May 23-26 2000.
- [BSL⁺96] Andreas Buja, Deborah F. Swayne, Michael L. Littman, Nathaniel Dean, and Heike Hofmann. Interactive Data Visualization with Multidimensional Scaling. *Journal of Computational and Graphical Statistics*, 5:78–99, 1996.
- [BYRN99] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison Wesley, 1999.
- [CCH06] Òscar Celma, Pedro Cano, and Perfecto Herrera. Search Sounds: An Audio Crawler Focused on Weblogs. In *Proceedings of the 7th International Society for Music Information Retrieval Conference (ISMIR)*, Victoria, Canada, 2006.

- [CF03] Matthew Cooper and Jonathan Foote. Summarizing Popular Music via Structural Similarity Analysis. In *Proceedings IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, October 2003.
- [Che53] E. C. Cherry. Some Experiments on the Recognition of Speech with One and Two Ears. *Journal of the Acoustical Society of America*, 25:975–979, 1953.
- [Che01] Elaine Chew. Modeling Tonality: Applications to Music Cognition. In Johanna D. Moore and Keith Stenning, editors, *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*, pages 206–211, Edinburgh, Scotland, UK, August 1-4 2001. Lawrence Erlbaum Assoc. Pub, Mahwah, NJ/London.
- [CJJ04] Sally J. Cunningham, Matt Jones, and Steve Jones. Organizing Digital Music for Use: An Examination of Personal Music Collections. In *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR)*, Barcelona, Spain, 2004.
- [CKGB02] Pedro Cano, Martin Kaltenbrunner, Fabien Gouyon, and Eloi Batlle. On the Use of FastMap for Audio Retrieval and Browsing. In *Proceedings of the 3rd International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France, 2002.
- [CL07] Oscar Celma and Paul Lamere. Music Recommendation Tutorial [online]. 2007. zugreifbar unter: <http://www.iaa.upf.edu/~ocelma/MusicRecommendationTutorial-ISMIR2007/> [letzter Abruf 15. Juni 2009].
- [CML⁺02] L. P. Clarisse, J. P. Martens, M. Lesaffre, B. De Baets, H. De Mayer, and M. Leman. An Auditory Model Based Transcriber of Singing Sequences. In *Proceedings of the 3rd International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France, 2002.
- [CR74] E. Catmull and R. Rom. A Class of Local Interpolating Splines. *Computer Aided Geometric Design*, pages 317–326, 1974.
- [CRB03] Sally J. Cunningham, Nina Reeves, and Matthew Britland. An Ethnographic Study of Music Information Seeking: Implications for the Design of a Music Digital Library. In *JCDL '03: Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries*, pages 5–16. IEEE Computer Society, 2003.
- [CT65] J.W. Cooley and J.W. Tukey. An Algorithm for the Machine Calculation of Complex Fourier Series. *Math. Computation*, 19:297–301, April 1965.
- [CV03] Wei Chai and Barry Vercoe. Structural Analysis of Musical Signals for Indexing and Thumbnailing. In *ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, Houston, Texas, USA, May 2003.
- [DHS01] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley & Sons, 2nd edition, 2001.
- [Dix01] Simon Dixon. Automatic Extraction of Tempo and Beat from Expressive Performances. *Journal of New Music Research*, 30:39–58, 2001.

- [dL77] J. de Leeuw. Applications of Convex Analysis to Multidimensional Scaling. *Recent Developments in Statistics*, pages 133–145, 1977.
- [dL01] J. de Leeuw. Multidimensional Scaling. *International Encyclopedia of the Social and Behavioral Sciences*, 2001.
- [DMR00] Michael Dittenbach, Dieter Merkl, and Andreas Rauber. The Growing Hierarchical Self-Organizing Map. *Neural Networks, IEEE - INNS - ENNS International Joint Conference on*, 6:6015, 2000.
- [DPW03] Simon Dixon, Elias Pampalk, and Gerhard Widmer. Classification of Dance Music by Periodicity Patterns. In *Proceedings of the 4th International Society for Music Information Retrieval Conference (ISMIR)*, Baltimore, MD, 2003.
- [DR03] Shyamala Doraisamy and Stefan Rüger. Robust Music Retrieval With N-Grams. *Journal of Intelligent Information Systems*, 21(1):53–70, 2003.
- [EW02] Daniel P. W. Ellis and Brian Whitman. The Quest for Ground Truth in Musical Artist Similarity. In *Proceedings of the 3rd International Society for Music Information Retrieval Conference (ISMIR)*, pages 170–177, Paris, France, 2002.
- [FB01] Mikael Fernström and Eoin Brazil. Sonic Browsing: An Auditory Tool for Multimedia Asset Management. In *Proceedings of the 2001 International Conference on Auditory Display*, Espoo, Finland, 2001.
- [FCN02] Jonathan Foote, Matthew Cooper, and Unjung Nam. Audio Retrieval by Rhythmic Similarity. In *Proceedings of the 3rd International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France, 2002.
- [FL95] C. Faloutsos and K. Lin. FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. In *Proceedings of the 1995 ACM SIGMOD*, pages 163–174, 1995.
- [Foo97] J.T. Foote. Content-based Retrieval of Music and Audio. *SPIE*, pages 138–147, 1997.
- [FU01] Jonathan Foote and Shingo Uchihashi. The Beat Spectrum: A New Approach to Rhythm Analysis. In *Proceedings IEEE International Conference on Multimedia and Expo (ICME)*, 2001.
- [GFW08] Martin Gasser, Arthur Flexer, and Gerhard Widmer. Streamcatcher: Integrated Visualization of Music Clips and Online Audio Streams. In *Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR)*, Philadelphia, 2008.
- [GG05] Masataka Goto and Takayuki Goto. Musicream: New Music Playback Interface for Streaming, Sticking, Sorting, and Recalling Musical Pieces. In *Proceedings of the 6th International Society for Music Information Retrieval Conference (ISMIR)*, London, England, 2005.

- [GLCS95] Asif Ghias, Jonathan Logan, David Chamberlin, and Brian C. Smith. Query by Humming: Musical Information Retrieval in an Audio Database. *ACM Multimedia*, pages 231–236, 1995.
- [Goo01] Michael Good. MusicXML for Notation and Analysis. In Walter B. Hewlett and Eleanor Selfridge-Field, editors, *The Virtual Score: Representation, Retrieval, Restoration*, pages 113–124, Cambridge, 2001. MIT Press.
- [Got03] M. Goto. A Chorus-Section Detecting Method For Musical Audio Signals. In *Proceeding IEEE ICASSP*, pages 437 – 440, 2003.
- [GWSK06] William T. Glaser, Timothy B. Westergren, Jeffrey P. Stearns, and Jonathan M. Kraft. Consumer Item Matching Method and System. *US Patent Number 7003515*, 2006.
- [HAE03] J. Herre, E. Allamanche, and C. Ertel. How similar Do Songs Sound? Towards Modeling Human Perception of Musical Similarity. In *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Mohonk, NY (USA), October 2003.
- [HDL⁺08] Xiao Hu, J. Stephen Downie, Cyril Laurier, Mert Bay, and Andrea F. Ehmann. The 2007 MIREX Audio Classification Task: Lessons Learned. In *Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR)*, Philadelphia, 2008.
- [HGR06] Amaury Hazan, Maarten Grachten, and Rafael Ramirez. Evolving Performance Models by Performance Similarity: Beyond Note-to-note Transformations. In *Proceedings of the 7th International Society for Music Information Retrieval Conference (ISMIR)*, Victoria, Canada, 2006.
- [HIMS07] Keiichiro Hoashi, Hiromi Ishizaki, Kazunori Matsumoto, and Fumiaki Sugaya. Content-based Music Retrieval Using Query Integration for users with Diverse Preferences. In *Proceedings of the 8th International Society for Music Information Retrieval Conference (ISMIR)*, Vienna, Austria, 2007.
- [HL06] Masatoshi Hamanaka and Seunghee Lee. Music Scope Headphones: Natural User Interface for Selection of Music. In *Proceedings of the 7th International Society for Music Information Retrieval Conference (ISMIR)*, Victoria, Canada, 2006.
- [HMI03] Keiichiro Hoashi, Kazunori Matsumoto, and Naomi Inoue. Personalization of User Profiles for Content-based Music Retrieval Based on Relevance Feedback. In *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, pages 110–119, New York, NY, USA, 2003. ACM.
- [HMT07] Stephen Hitchner, Jennifer Murdoch, and George Tzanetakis. Music Browsing Using a Tabletop Display. In *Proceedings of the 8th International Society for Music Information Retrieval Conference (ISMIR)*, Vienna, Austria, 2007.
- [Hof99] Thomas Hofmann. Probabilistic Latent Semantic Analysis. In *In Proc. of Uncertainty in Artificial Intelligence, UAI'99*, pages 289–296, 1999.

- [Hon03] J. Hong. Investigation Expressive Timing and Dynamics in Recorded Cello. *Psychology of Music*, 31(3):340 – 352, 2003.
- [Hon04a] Seung-Hyun Hong. The Effect of Digital Technology on the Sales of Copyrighted Goods: Evidence from Napster. Working Paper. Department of Economics, Stanford University, 2004.
- [Hon04b] Seung-Hyun Hong. The Effect of Napster on Recorded Music Sales: Evidence from the Consumer Expenditure Survey. SIEPR Discussion Paper No. 03-18. Stanford University, Januar 2004.
- [HS72] Peter Hanser-Strecker. Plagiate. In *Schlager in Deutschland*. Breitkopf & Härtel, 1972.
- [HS08] Andre Holzapfel and Yannis Stylianou. Beat Tracking Using Group Delay Based Onset Detection. In *Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR)*, Philadelphia, 2008.
- [Hum] The Humdrum Toolkit: Software for Music Research [online]. zugreifbar unter: <http://music-cog.ohio-state.edu/Humdrum> [letzter Abruf 15. März 2009].
- [Hur97] D. B. Huron. *Humdrum and Kern: Selective Feature Encoding*, pages 375 – 401. MIT Press, Cambridge, MA, 1997.
- [IAD⁺92] Y. Idan, J.M. Auger, N. Darbel, M. Sales, R. Chevallier, B. Dorizzi, and G. Cazuguel. Comparative Study of Neural Networks and Non-Parametric Statistical Methods for Off-line Handwritten Character Recognition. In *Proceedings ICANN'92*, Brighton, 1992.
- [IFP07] IFPI. Recorded Music Sales [online]. 1995–2007. zugreifbar unter: <http://www.ifpi.org/content/library/worldsales.pdf> [letzter Abruf 14. März 2009].
- [IFP09] IFPI. Digital music report 2009 [online]. Januar 2009. zugreifbar unter: <http://www.ifpi.org/content/library/DMR2009.pdf> [letzter Abruf 14. März 2009].
- [IM92] Luc M. Van Immerseel and Jean-Pierre Martens. Pitch and Voiced/Unvoiced Determination With an Auditory Model. *Journal of the Acoustical Society of America*, 91(6):3511–3526, June 1992.
- [Jac06] Kurt Jacobson. A Multifaceted Approach to Music Similarity. In *Proceedings of the 7th International Society for Music Information Retrieval Conference (ISMIR)*, Victoria, Canada, 2006.
- [JECJ07] Jesper Jøjvang Jensen, Daniel P.W. Ellis, Mads G. Christensen, and Søren Holdt Jensen. Evaluation of Distance Measures between Gaussian Mixture Models of MFCCs. In *Proceedings of the 8th International Society for Music Information Retrieval Conference (ISMIR)*, Vienna, Austria, 2007.
- [Jen99] Kristoffer Jensen. *Timbre Models of Musical Sounds*. PhD thesis, University of Copenhagen, 1999.

- [JFB02] P.N. Juslin, A. Frieberg, and R. Bresing. Toward a Computational Model of Expression in Performance: The GERM model. *Musicae Scientiae*, pages 63–122, 2001–2002.
- [JLZ⁺02] D.-N. Jiang, L. Lu, H.-J. Zhang, J.-H. Tao, and L.-H. Cai. Music Type Classification by Spectral Contrast Feature. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, Lausanne (Switzerland), August 2002.
- [JN84] N. S. Jayant and P. Noll. *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice-Hall, Englewood Cliffs, NJ, 1984.
- [JOPP98] Murray Jennex, Lorne Olfman, Pituma Panthawi, and Yong-Tae Park. An Organizational Memory Information Systems Success Model: An Extension of DeLone and McLean’s I/S Success Model . In *IEEE 31st Annual Hawaii International Conference on System Sciences (1)*, pages 157–165, 1998.
- [Jus00] P.N. Juslin. Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception. *Journal of Experimental Psychology: Human Perception and Performance*, 26(6), 2000.
- [Kei02] Daniel A. Keim. Datenvisualisierung und Data Mining. *Datenbank-Spektrum*, 2, 2002.
- [Kel07] Thorben Keller. Untersuchungen zur Anpassung von inhaltsbasierten an Community-basierte Ähnlichkeitsmaße für Musik. Master’s thesis, RWTH University, Aachen, 2007.
- [Kla04] Anssi P. Klapuri. Automatic Music Transcription As We Know It Today. *Journal of New Music Research*, 33(3):269–282, 2004.
- [KMW94] J. Krimphoff, S. McAdams, and S. Winsberg. Caractérisation du timbre des sons complexes. II : Analyses acoustiques et quantification psychophysique. *Journal de Physique*, 4:625–628, 1994.
- [Koh82] T. Kohonen. Self-Organizing Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43:59–69, 1982.
- [Koh92] T. Kohonen. New Developments of Learning Vector Quantization and the Self-Organizing Map. In *SYNAPSE’92, Symposium on Neural Networks*, Osaka, Japan, 1992.
- [Koh01] T. Kohonen. *Self-Organizing Maps*. Springer, Berlin, 2001.
- [Kor01] Andreas Kornstädt. The JRing System for Computer-Assisted Musicological Analysis. In *Proceedings of the 2nd International Society for Music Information Retrieval Conference (ISMIR)*, Bloomington, Indiana, USA, 2001.
- [KOS⁺96] T. Kohonen, E. Oja, O. Simula, A. Visa, and J. Kangas. Engineering Applications of the Self-Organizing Map. *Proceedings of the IEEE*, 84(10):1356–1384, 1996.

- [KPS⁺08] Peter Knees, Tim Pohle, Markus Schedl, Dominik Schnitzer, and Klaus Seyerlehner. A Document-centered Approach to a Natural Language Music Search Engine. In *Proceedings of the 30th European Conference on Information Retrieval (ECIR'08)*, Glasgow, Scotland, UK, March 30-April 3 2008.
- [KPSW07a] Peter Knees, Tim Pohle, Markus Schedl, and Gerhard Widmer. A Music Search Engine Built upon Audio-based and Web-based Similarity Measures. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'07)*, Amsterdam, The Netherlands, July 23-27 2007.
- [KPSW07b] Peter Knees, Tim Pohle, Markus Schedl, and Gerhard Widmer. Exploring Music Collections in Virtual Landscapes. *IEEE MultiMedia*, 14(3):46–54, 2007.
- [Kru64a] Joseph B. Kruskal. Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika*, 29:1–27, 1964.
- [Kru64b] Joseph B. Kruskal. Nonmetric Multidimensional Scaling: A Numerical Method. *Psychometrika*, 29:115–129, 1964.
- [Kru90] Carol L. Krumhansl. *Cognitive Foundations of Musical Pitch*. Oxford University Press, New York, 1990.
- [KS97] T. Kohonen and P. Somervuo. Self-Organizing Maps of Symbol String with Application to Speech Recognition. In *Proceedings of the Workshop on Self-Organizing Maps (WSOM'97)*, pages 2–7, Espoo, Finland, 1997.
- [KS03] Victor Kulesh and Ishwar Sethi. Indexing and Retrieval of Musica via Gaussian Mixture Models. In *Proceedings of the 3rd International Workshop on Content-Based Multimedia Indexing (CBMI)*, Rennes (France), 2003.
- [Kut04] Heinrich Kuttruff. *Akustik: Eine Einführung*. Hirzel, Stuttgart, 2004.
- [KW07] Peter Knees and Gerhard Widmer. Searching for Music Using Natural Language Queries and Relevance Feedback. In *Proceedings of the 5th International Workshop on Adaptive Multimedia Retrieval (AMR'07)*, Paris, France, July 2007.
- [Lar08] Olivier Lartillot. MIRtoolbox 1.1 User's Manual. Technical report, Finnish Center of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Finland, 2008.
- [LC00] Beth Logan and Stephen Chu. Music Summarization Using Key Phrases. In *Proceedings IEEE ICASSP*, pages 749–752, 2000.
- [LD06] Audrey Laplante and J. Stephen Downie. Everyday Life Music Information-Seeking Behaviour of Young Adults. In *Proceedings of the 7th International Society for Music Information Retrieval Conference (ISMIR)*, Victoria, Canada, 2006.
- [LdMA02] R. López de Mántaras and J.L. Arcos. From Composition to Expressive Performances. *AI Magazine*, 23(3):43–57, 2002.

- [LE07] Paul Lamere and Douglas Eck. Using 3D Visualizations to Explore and Discover Music. In *Proceedings of the 8th International Society for Music Information Retrieval Conference (ISMIR)*, Vienna, Austria, 2007.
- [LEB03] Beth Logan, Daniel P. W. Ellis, and Adam Berenzweig. Toward Evaluation Techniques for Music Similarity. *IEEE ICME*, 2003.
- [LeC09] Jennifer LeClaire. YouTube Negotiating to Become a Digital MTV [online]. 2009. zugreifbar unter: http://www.sci-tech-today.com/news/YouTube-Could-Become-Digital-MTV/story.xhtml?story_id=011001CEH0R9&full_skip=1 [letzter Abruf 14. März 2009].
- [Lib07] Libraryman. Pandora Thieves Qualities [online]. 2007. zugreifbar unter: <http://www.flickr.com/photos/libraryman/1225285863/> [letzter Abruf 11. Juni 2009].
- [LJ09] Dominik Lübbers and Matthias Jarke. Adaptive Multimodal Exploration of Music Collections. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, 2009.
- [LKM04] Beth Logan, Andrew Kositsky, and Pedro Moreno. Semantic Analysis of Song Lyrics. In *Proceedings IEEE International Conference on Multimedia and Expo (ICME)*, Taipei, Taiwan, June 2004.
- [Log00] Beth Logan. Mel Frequency Cepstral Coefficients for Music Modeling. In *Proceedings of International Symposium for Music Information Retrieval (ISMIR)*, Plymouth, Massachusetts (USA), 2000.
- [LOL03] Tao Li, Mitsunori Ogihara, and Qi Li. A Comparative Study on Content-Based Music Genre Classification. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 282–289, Toronto, Canada, 2003.
- [LS01] Beth Logan and Ariel Salomon. A Music Similarity Function Based on Signal Analysis. In *Proceedings of the IEEE International Conference on Multimedia Expo (ICME)*, pages 745 – 748, 2001.
- [LS07] Mark Levy and Mark Sandler. A Semantic Space for Music Derived From Social Tags. In *Proceedings of the 8th International Society for Music Information Retrieval Conference (ISMIR)*, Vienna, Austria, 2007.
- [LSDM01] Dongge Li, Ishwar K. Sethi, Nevenka Dimitrova, and Tom McGee. Classification of General Audio Data for Content-Based Retrieval. *Pattern Recognition Letters*, 22:533–544, 2001.
- [LST04] Aristomenis S. Lampropoulos, Dionisios N. Sotiropoulos, and George A. Tsihrintzis. Individualization of Music Similarity Perception via Feature Subset Selection. In *IEEE International Conference on Systems, Man and Cybernetics*, pages 552 – 556, 2004.

- [LSY03] Greg Linden, Brent Smith, and Jeremy York. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing*, 7(1):76 – 80, January 2003.
- [LT07] Olivier Lartillot and Petri Toiviainen. A Matlab Toolbox for Musical Feature Extraction from Audio. In *Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07)*, 2007.
- [Lü05] Dominik Lübbers. soniXplorer: Combining Visualization and Auralization for Content-based Exploration of Music Collections. In *Proceedings of the 6th International Society for Music Information Retrieval Conference (ISMIR)*, London, England, 2005.
- [MB03] Martin F. McKinney and Jeroen Breebaart. Features for Audio and Music Classification. In *Proceedings of the 4th International Society for Music Information Retrieval Conference (ISMIR)*, Baltimore, MD, 2003.
- [Mer03] D. Meredith. Pitch Spelling Algorithms. In *Proceedings of the 5th Triennial ESCOM Conference*, Hannover, Germany, 2003.
- [Mey00] Carsten Meyer. Doppelt blind - MP3 gegen CD: Der Hörtest. *Computer und Technik c't*, 3:144, 2000.
- [Mic04] Robert Michel. The Impact of the Digital Age on the Music Industry: An Empirical Analysis. Working Paper, 2004.
- [Mit97] Tom M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.
- [MK07] Meinard Müller and Frank Kurth. Towards structural analysis of audio recordings in the presence of musical variations. *EURASIP J. Appl. Signal Process.*, 2007(1):163–163, 2007.
- [MPE06] Michael I. Mandel, Graham E. Poliner, and Daniel P.W. Ellis. Support Vector Machine Active Learning for Music Retrieval. *Multimedia Systems*, 12(1):3–13, 2006.
- [MRC03] Alistair Morrison, Greg Ross, and Matthew Chalmers. Fast Multidimensional Scaling through Sampling, Springs and Interpolation. *Information Visualization*, 2(1):68–77, 2003.
- [MS88] Gary Machionini and Ben Shneiderman. Finding Facts vs. Browsing : Knowledge in Hypertext Systems. *Computer*, 21(1):70 – 80, 1988.
- [MS99] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.
- [MS08a] Terence Magno and Carl Sable. A Comparison of Signal-Based Music Recommendation to Genre Labels, Collaborative Filtering, Musicological Analysis, Human Recommendation, and Random Baseline. In *Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR)*, Philadelphia, 2008.

- [MS08b] Sarah McBride and Ethan Smith. Music Industry to Abandon Mass Suits. The Wall Street Journal, 19. Dezember 2008. zugreifbar unter: http://online.wsj.com/article/SB122966038836021137.html?mod=googlenews_wsj [letzter Abruf 14. März 2009].
- [MT06] Jennifer Murdoch and George Tzanetakis. Interactive Content-Aware Music Browsing using the Radio Drum. In *Proceedings IEEE International Conference on Multimedia and Expo*, pages 937–940, Los Alamitos, CA, USA, 2006. IEEE Computer Society.
- [MUNS05] Fabian Mörchen, Alfred Ultsch, Mario Nöcker, and Christian Stamm. Data-bionic Visualization of Music Collections According to Perceptual Distance. In *Proceedings of the 6th International Society for Music Information Retrieval Conference (ISMIR)*, London, England, 2005.
- [MUTL06] Fabian Mörchen, Alfred Ultsch, Michael Thies, and Ingo Löhken. Modeling Timbre Distance With Temporal Statics From Polyphonic Music. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1), January 2006.
- [NDR05] Robert Neumayer, Michael Dittenbach, and Andreas Rauber. PlaySOM and PocketSOMPlayer - Alternative Interfaces to Large Music Collections. In *Proceedings of the 6th International Society for Music Information Retrieval Conference (ISMIR)*, London, England, 2005.
- [Nei07] Jan Neis. Interaktive multimodale Exploration von Musiksammlungen mittels auralisierter Merkmalskarten. Master's thesis, RWTH Aachen University, 2007.
- [NK02] Andreas Nürnberger and Aljoscha Klose. Improving Clustering and Visualization of Multimedia Data Using Interactive User Feedback. In *Proc. of the 9th International Conference on Information Processin and Management of Uncertainty in Knowledge-Based Systems (IPMU 2002)*, pages 993–999, 2002.
- [NS08] Andreas Nürnberger and Sebastian Stober. User Modelling for Interactive User-Adaptive Collection Structuring. *Adaptive Multimedial Retrieval: Retrieval, User, and Semantics: 5th International Workshop, AMR 2007, Paris, France, July 5-6, 2007 Revised Selected Papers*, pages 95–108, 2008.
- [Nü01] Andreas Nürnberger. Interactive Text Retrieval Supported by Growing Self-Organizing Maps. In *Proc. of the International Workshop on Information Retrieval (IR 2001)*, pages 61–70, Infotech, Oulu, Finland, 2001.
- [OKS06] Nuria Oliver and Lucas Kreger-Stickles. PAPA: Physiology and Purpose-Aware Automatic Playlist Generation. In *Proceedings of the 7th International Society for Music Information Retrieval Conference (ISMIR)*, Victoria, Canada, 2006.
- [OS07] Felix Oberholzer and Koleman Strumpf. The Effect of File Sharing on Record Sales: An Empirical Analysis. *Journal of Political Economy*, 115(1), 2007.
- [OSB99] Alan V. Oppenheim, Ronald W. Schafer, and John R. Buck. *Discrete-Time Signal Processing*. Prentice-Hall, 2nd edition, 1999.

- [Par04] Bryan Pardo. Tempo Tracking with a Single Oscillator. In *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR)*, Barcelona, Spain, 2004.
- [Pau03] Steffen Pauws. Effect of Song Familiarity, Singing Training and Recent Song Exposure on the Singing of Melodies. In *Proceedings of the 4th International Society for Music Information Retrieval Conference (ISMIR)*, pages 57–64, Baltimore, MD, 2003.
- [PBR02] Geoffroy Peeters, Amaury La Burthe, and Xavier Rodet. Toward Automatic Music Audio Summary Generation from Signal Analysis. In *Proceedings of the 3rd International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France, 2002.
- [PC00] François Pachet and Daniel Cazaly. A Taxonomy of Musical Genres. In *Content-Based Multimedia Information Access Conference (RIAO)*, Paris, 2000.
- [PDW03a] Elias Pampalk, Simon Dixon, and Gerhard Widmer. Exploring Music Collections by Browsing Different Views. In *Proceedings of the 4th International Society for Music Information Retrieval Conference (ISMIR)*, pages 201 – 208, Baltimore, MD, 2003.
- [PDW03b] Elias Pampalk, Simon Dixon, and Gerhard Widmer. On the Evaluation of Perceptual Similarity Measures for Music. In *Proc. of the 6th Int. Conference on Digital Audio Effects (DAFx-03)*, London, UK, September 2003.
- [PG06] Elias Pampalk and Masataka Goto. MusicRainbow: A New user Interface to Discover Artists Using Audio-based Similarity and Web-based Labeling. In *Proceedings of the 7th International Society for Music Information Retrieval Conference (ISMIR)*, Victoria, Canada, 2006.
- [PG07] Elias Pampalk and Mastaka Goto. Musicsun: A New Approach to Artist Recommendation. In *Proceedings of the 8th International Society for Music Information Retrieval Conference (ISMIR)*, Vienna, Austria, 2007.
- [PH01] Klaus Poeck and Werner Hacke. *Neurologie*. Springer, Berlin, 11. edition, 2001.
- [Pho75] Bui Tuong Phong. Illumination for Computer Generated Pictures. *Communications of the ACM*, 18(6):311–317, 1975.
- [Pic04] Jeremy Pickens. *Harmonic Modeling for Polyphonic Music Retrieval*. PhD thesis, University of Massachusetts Amherst, 2004.
- [PK02] Jouni Paulus and Anssi Klapuri. Measuring the Similarity of Rhythmic Patterns. In *Proceedings of the 3rd International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France, 2002.
- [PK06] Jouni Paulus and Anssi Klapuri. Music Structure Analysis by Finding Repeated Parts. In *Proceedings Audio and Music Computing for Multimedia Workshop (AMCMM)*, Santa Barbara, California, USA, October 2006.

- [PL01] François Pachet and Damien Laigre. A Naturalist Approach to Music File Name Analysis. In *Proceedings of the 2nd International Society for Music Information Retrieval Conference (ISMIR)*, Bloomington, Indiana, USA, 2001.
- [PPW05] T. Pohle, E. Pampalk, and G. Widmer. Evaluation of Frequently Used Audio Features for Classification of Music Into Perceptual Categories. In *Proceedings of the 4th International Workshop on Content-Based Multimedia Indexing (CBMI 05)*, Riga, Latvia, June 2005.
- [PRM02] Elias Pampalk, Andreas Rauber, and Dieter Merkl. Content-based Organization and Visualization of Music Archives. In *Proceedings of the ACM Multimedia Conference (ACM MM02)*, pages 570 – 579, 2002.
- [PW04] Martin Peitz and Patrick Waelbroeck. The Effect of Internet Piracy on Music Sales: Cross-Section Evidence. *Review of Economic Research on Copyright Issues*, 1:71–79, 2004.
- [RB81] B. Lindblom R. Bladon. Modeling the Judgement of Vowel Quality Differences. *Journal of Acoustic Society America*, 69(5):1414–1422, 1981.
- [RH00] Yong Rui and Thomas Huang. Optimizing Learning in Image Retrieval. In *IEEE Conf. on CVPR*, pages 236–243, 2000.
- [Rif02] R. M. Rifkin. *Everything Old is New Again: A Fresh Look at Historical Approaches to Machine Learning*. PhD thesis, Massachusetts Institute of Technology, 2002.
- [RJ93] L. Rabiner and B. Juang. *Fundamentals of Speech Recognition*. Prentice-Hall, 1993.
- [RMD02] Andreas Rauber, Dieter Merkl, and Michael Dittenbach. The Growing Hierarchical Self-Organizing Map: Exploratory Analysis of High-Dimensional Data. *IEEE Transactions on Neural Networks*, 13:1331–1341, 2002.
- [Roa96] Curtis Roads. *The Computer Music Tutorial*. MIT Press, 1996.
- [Roc71] J.J. Rocchio. *Relevance Feedback in Information Retrieval in The SMART Retrieval System: Experiments in Automatic Document Processing*, chapter 14. Prentice Hall Inc., 1971.
- [RPM02] Andreas Rauber, Elias Pampalk, and Dieter Merkl. Using Psycho-Acoustic Models and Self-Organizing Maps to Create a Hierarchical Structuring of Music by Musical Styles. In *Proceedings of the 3rd International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France, 2002.
- [RTG98] Y. Rubner, C. Tomasi, and L. Guibas. The Earth Mover’s Distance as a Metric for Image Retrieval. Technical report, Stanford University, 1998.
- [RW06] Rafael Rob and Joel Waldfogel. Piracy on the High C’s: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students. *Journal of Law and Economics*, 49:29–62, 2006.

- [SAH79] M.R. Schroeder, B.S. Atal, and J.L. Hall. Optimizing Digital Speech Coders by Exploiting Masking Properties of the Human Ear. *Journal of Acoustic Society America*, 66(6):1647–1652, 1979.
- [SB88] G. Salton and C. Buckley. Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing and Management*, 24(5):513–523, 1988.
- [SCBG08] Mohamed Sordo, Òscar Celma, Martýn Blech, and Enric Guaus. The Quest for Musical Genres: Do the Experts and the Wisdom of Crowds Agree? In *Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR)*, Philadelphia, 2008.
- [She62] R.N. Shepard. The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function. *Psychometrika*, 27, 1962.
- [SK05] Richard Stenzel and Thomas Kamps. Vergleiche und Kombinationen von Systemen zur Ähnlichkeitsbestimmung für Musik. *Informatik Spektrum*, 15:375 – 380, Oktober 2005.
- [SKKR01] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-Based Collaborative Filtering Recommendation Algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295, 2001.
- [SLC07] Mohamed Sordo, Cyril Laurier, and Oscar Celma. Annotating Music Collections: How Content-Based Similarity Helps to Propagate Labels. In *Proceedings of the 8th International Society for Music Information Retrieval Conference (ISMIR)*, Vienna, Austria, 2007.
- [SLT07] D.N. Sotiropoulos, A.S. Lampropoulos, and G.A. Tsihrintzis. Evaluation of Modeling Music Similarity Perception Via Feature Subset Selection. In *Proceedings of the 11th international conference on User Modeling*, pages 288 – 297, 2007.
- [SLT08] Dionysios N. Sotiropoulos, Aristomenis S. Lampropoulos, and George A. Tsihrintzis. MUSIPER: a system for modeling music similarity perception based on objective feature subset selection. *User Modeling and User-Adapted Interaction*, 18(4):315–348, 2008.
- [SN09] Sebastian Stober and Andreas Nürnberger. User-Adaptive Music Information Retrieval. *Künstliche Intelligenz*, 2:54–57, 2009.
- [SRU04] Joshua Stoddard, Christopher Raphael, and Paul E. Utgoff. Well-Tempered Spelling: A Key-Invariant Pitch Spelling Algorithm. In *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR)*, pages 106 – 111, Barcelona, Spain, 2004.
- [S XK04] Xi Shao, Changsheng Xu, and Mohan S. Kankanhalli. Unsupervised Classification of Music Genre Using Hidden Markov Model. In *Proceedings IEEE International Conference on Multimedia and Expo ICME 04*, volume 3, pages 2023–2026, 2004.

- [SZM06] N. Scaringella, G. Zoia, and D. Mlynek. Automatic genre classification of music content: a survey. *Signal Processing Magazine, IEEE*, 23(2):133 – 141, March 2006.
- [TBL08] Douglas Turnbull, Luke Barrington, and Gert Lanckriet. Five Approaches to Collecting Tags for Music. In *Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR)*, Philadelphia, 2008.
- [TBTL07] Douglas Turnbull, Luke Barrington, David Torres, and Gert Lanckriet. Towards Musical Query-by-semantic-description Using the CAL500 Data Set. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 439–446, 2007.
- [TBTL08] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. Semantic Annotation and Retrieval of Music and Sound Effects. *IEEE Transactions on Audio, Speech and Language Processing*, 16(2), 2008.
- [TC00] George Tzanetakis and Perry Cook. MARSYAS: A Framework for Audio Analysis. *Organized Sound*, 4(3), 2000.
- [TC01] George Tzanetakis and Perry Cook. Marsyas3D: A Prototype Audio Browser-Editor Using a Large Scale Immersive Visual and Audio Display. In *Proceedings International Conference on Auditory Display*, Espoo, Finland, 2001.
- [TC02] George Tzanetakis and Perry Cook. Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing*, 2002.
- [Tem99] David Temperley. What’s Key for Key? The Krumhansl-Schmuckler Key Finding Algorithm Reconsidered. *Music Perception*, 17(1):65 – 100, 1999.
- [Tem02] David Temperley. A Bayesian Approach to Key-Finding. *Music and Artificial Intelligence*, pages 195 – 206, 2002.
- [Ter79] E. Terhardt. Calculating Virtual Pitch. *Hearing Research*, 1:155–182, 1979.
- [TGV⁺03] Rainer Typke, Panos Giannopoulos, Remco C. Veltkamp, Frans Wiering, and René van Oostrum. Using Transportation Distances for Measuring Melodic Similarity. In *Proceedings of the 4th International Society for Music Information Retrieval Conference (ISMIR)*, Baltimore, MD, 2003.
- [THA04] Marc Torrens, Patrick Hertzog, and Josep-Lluís Arcos. Visualizing and Exploring Personal Music Libraries. In *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR)*, Barcelona, Spain, 2004.
- [Tor52] Warren S. Torgerson. Multidimensional Scaling: I. Theory and Method. *Psychometrika*, pages 401–419, 1952.
- [TTBL07] David Torres, Douglas Turnbull, Luke Barrington, and Gert Lanckriet. Identifying Words that are Musically Meaningful. In *Proceedings of the 8th International Society for Music Information Retrieval Conference (ISMIR)*, Vienna, Austria, 2007.

- [TWV04] Rainer Typke, Frans Wiering, and Remco C. Veltkamp. A Search Method for Notated Polyphonic Music with Pitch and Tempo Fluctuations. In *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR)*, Barcelona, Spain, 2004.
- [Tza02] George Tzanetakis. *Manipulation, Analysis and Retrieval Systems for Audio Signals*. PhD thesis, Princeton University, 2002.
- [UJ06] Jana Urban and Joemon M. Jose. EGO: A Personalised Multimedia Management Tool. *International Journal of Intelligent Systems*, 21(7):725–745, 2006.
- [Ult92] Alfred Ultsch. Self-organizing Neural Networks for Visualization and Classification. In *Proc. Conf. Soc. for Information and Classification*, Dortmund, Germany, April 1992.
- [Ult03a] Alfred Ultsch. Maps for the Visualization of High Dimensional Data Spaces. In *Proc. WSOM '03*, Japan, 2003.
- [Ult03b] Alfred Ultsch. Pareto Density Estimation: Probability Density Estimation for Knowledge Discovery. In *Proceedings GfKI*, Cottbus, Germany, 2003.
- [Vap98] V.N. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, 1998.
- [VB04] Shankar Vembu and Stephan Baumann. A Self-Organizing Map Based Knowledge Discovery for Music Recommendation Systems. In *Proceedings of the 2nd International Symposium on Computer Music Modeling and Retrieval*, Esbjerg, Denmark, 2004.
- [vGVvdW04] Rob van Gulik, Fabio Vignoli, and Huub van de Wetering. Mapping Music in the Palm of Your Hand, Explore and Discover Your Collection. In *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR)*, Barcelona, Spain, 2004.
- [VKE03] T. Viitaniemi, A. Klapuri, and A. Eronen. A Probabilistic Model for the Transcription of Single-Voice Melodies. In *Proc. Finnish Signal Processing Symposium*, Tampere, Finland, 2003.
- [VP05] Fabio Vignoli and Steffen Pauws. A Music Retrieval System Based on User-Driven Similarity and Its Evaluation. In *Proceedings of the 6th International Society for Music Information Retrieval Conference (ISMIR)*, London, England, 2005.
- [vSDR01] Thomas von Schroeter, Shyamala Doraisamy, and Stefan M. Rüger. From Raw Polyphonic Audio to Locating Recurring Themes. In *Proceedings of the 2nd International Society for Music Information Retrieval Conference (ISMIR)*, Bloomington, Indiana, USA, 2001.
- [WBKW96] E. Wold, T. Blum, D. Keislar, and J. Wheaton. Content-Based Classification Search and Retrieval of Audio. *IEEE Multimedia Magazine*, 3:27–36, July 1996.

- [WC04] Kris West and Stephen Cox. Feature and Classifiers for the Automatic Classification of Musical Audio Signals. In *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR)*, Barcelona, Spain, 2004.
- [WDG⁺03] Gerhard Widmer, Simon Dixon, Werner Goebl, Elias Pampalk, and Asmir Tobudic. In Search of the Horowitz Factor. In *Proceedings of the 5th International Conference on Discovery Science*, pages 13 – 21, 2003.
- [WE04] Brian Whitman and Daniel Ellis. Automatic Record Reviews. In *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR)*, Barcelona, Spain, 2004.
- [Wey02] Tilmann Weyde. Integrating Segmentation and Similarity in Melodic Analysis. In *Proceedings of the International Conference on Music Perception and Cognition*, pages 240 – 243, Sydney, Australia, 2002.
- [Wik09] Wikipedia. Music Genome Project [online]. 2009. zugreifbar unter: http://en.wikipedia.org/wiki/Music_Genome [letzter Abruf 11. Juni 2009].
- [WL02] Brian Whitman and Steve Lawrence. Inferring Descriptions and Similarity for Music from Community Metadata. In *Proceedings of the 3rd International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France, 2002.
- [WR02] Brian Whitman and Ryan Rifkin. Musical Query-by-Description as a Multiclass Learning Problem. In *Proceedings IEEE Multimedia Signal Processing Conference (MMSP)*, pages 153–156, 2002.
- [XMS⁺03] Changsheng Xu, Namunu C. Maddage, Xi Shao, Fang Cao, and Qi Tian. Musical Genre Classification Using Support Vector Machines. In *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 429–432, 2003.
- [XNJR03] Eric P. Xing, Andrew Y. Ng, Michael I. Jordan, and Stuart Russell. Distance Metric Learning, with Application to Clustering with Side-information. In *Advances in Neural Information Processing Systems 15*, pages 505–512. MIT Press, 2003.
- [Yan08] Qian Yang. Music Recommendation Based on Album Reviews. Master’s thesis, RWTH Aachen University, 2008.
- [Zen05] Alejandro Zentner. File Sharing and International Sales of Copyrighted Music: An Empirical Analysis with a Panel of Countries. *Topics in Economic Analysis & Policy*, 5(1), 2005.
- [Zen06] Alejandro Zentner. Measuring the Effect of Music Downloads on Music Purchases. *Journal of Law and Economics*, April 2006.
- [ZF99] E. Zwicker and H. Fastl. *Psychoacoustics, Facts and Models*. Springer, 2nd edition, 1999.
- [Zö02] Udo Zölzer, editor. *DAFX: Digital Audio Effects*. John Wiley & Sons, 2002.

Abbildungsverzeichnis

1.1	Verkauf physikalischer Tonträger	2
1.2	Einnahmen aus Online-Verkauf digitaler Musik weltweit	4
1.3	Invention 13 als Notentext	12
1.4	Invention 13 als Music-XML-Datei	13
1.5	Invention 13 als MIDI-Datei in Piano-Roll-Darstellung	16
1.6	Invention 13 als MIDI-Datei in Listen-Darstellung	17
1.7	Invention 13 in Wellenform-Darstellung	18
1.8	Invention 13 in Spektogramm-Darstellung	20
1.9	Screenshots des soniXplorer-Prototypen (I)	26
1.10	Screenshots des soniXplorer-Prototypen (II)	27
1.11	Kernprozesse der adaptiven multimodalen Exploration im soniXplorer	28
2.1	Prozessmodell Berechnung Klang-basierter Ähnlichkeit	33
2.2	Beat Histogramme unterschiedlicher Genres	42
2.3	Selbstähnlichkeitsmatrix nach Foote	45
2.4	Screenshot Erläuterung Pandora-Empfehlung	49
2.5	Tag-Cloud <i>Turn Me On</i> (Norah Jones)	54
2.6	IR-Ansatz Rezensions-basierte Alben-Empfehlung	61
2.7	Maschinelles Lernen zur Rezensions-basierte Alben-Empfehlung	62
2.8	Gewichtung von Musikähnlichkeits-Aspekten nach Baumann	67
2.9	Wurzel quadrierter Regressions-Fehler	71
2.10	Pearson Korrelationskoeffizient kollaborative und kombinierte inhaltsbasierte Ähnlichkeit	71
2.11	Gegenüberstellung Popularität und Rang von Last.fm Tags	75
3.1	Metadaten-basierte Visualisierung nach Torrens et al.	80
3.2	Sonic-Browser Visualisierung	81
3.3	Islands of Music-Visualisierung nach Pampalk	83

3.4	MusicMiner-Visualisierung	84
3.5	Kopfhörer-Sensoren	85
3.6	Musikkollektion als dreidimensionale Landschaft	87
3.7	Darstellungsmodi des <i>nepTune</i> -Systems	88
3.8	MDS-basierte dreidimensionale Kollektions-Visualisierung	96
3.9	Entfaltung somato-sensorischer Cortex	97
3.10	Rechteckige und hexagonale Netzwerktopologie	101
3.11	Regionen gleichen Abstandes vom zentralen Neuron	101
3.12	Geschlossene Karten-Topologien	102
3.13	Demonstrationskollektion auf zweidimensionaler SOM	106
3.14	Demonstrationskollektion im <i>soniXplorer</i>	107
3.15	Drahtmodell-Darstellung ohne Höhenprofil-Interpolation	108
3.16	Drahtmodell-Darstellung mit Höhenprofil-Interpolation	109
3.17	Terrain-Texturen	109
3.18	Texturiertes Terrain	110
3.19	Visualisierung von Stück-Eigenschaften	111
3.20	Erkennung von Liedern am Titel vs. Klang	112
3.21	Bildschirmkopie SonicRadar-Prototyp	114
3.22	Wiegabepegelverlauf $g_{PF}(\varphi)$	116
3.23	Visuelle Repräsentation einer Einschränkung des Wahrnehmungswinkels	116
3.24	Skizze zur optimierten Stückauswahl bei drei Titeln	118
3.25	<i>soniXplorer</i> im Solo-Modus	120
3.26	Xbox 360-Controller zur Steuerung des <i>soniXplorers</i>	121
3.27	<i>soniXplorer</i> -Blick auf eine Musikkollektion aus verschiedenen Höhen	122
4.1	Visuelle Repräsentation eines Stückes als Flügel-Symbol	124
4.2	Texturierte Schilder verschiedener Formen als Orientierungspunkte	126
4.3	Explorationsumgebung vor und nach Entfernung einer trennenden Bergkette	127
4.4	Aufbau einer zusammenhängenden trennenden Bergkette	128
5.1	Geschlechter-Verteilung in der Versuchsgruppe	150
5.2	Alters-Verteilung in der Versuchsgruppe	151
5.3	Anzahl Stunden Musikkonsum pro Woche	152
5.4	Anzahl CDs in persönlichem Besitz	152
5.5	Anzahl physikalischer Tonträger im Besitz der Probanden	153
5.6	Anzahl in Online-Shops heruntergeladener Stücke	153
5.7	Größe der MP3-Kollektionen der Probanden	154
5.8	Vertrautheit der Probanden mit Computerspielen	155
5.9	Reduktion der Suchzeit durch SOM-Strukturierung	156
5.10	Reduktion der Suchzeit bei akustischer Umgebung	157
5.11	Positionierungsfehler mit und ohne Anpassung	157
5.12	Relative Reduktion des Positionierungsfehlers durch Adaption	158

Lebenslauf

Persönliche Daten

Name: Dominik Ansgar Lübbers
Geburtsdatum: 04. Juli 1975
Geburstort: Osnabrück

Schulabschluss

Mai 1995 Allgemeine Hochschulreife (Abitur)
Gymnasium Marianum, Meppen

Wehrersatzdienst

1995 – 1996 Zivildienst im Krankenhaus Ludmillenstift, Meppen

Studium

1996 – 2002 Studium der Informatik an der RWTH Aachen
Anwendungsfach Medizin
Vertiefungsgebiet Informationssysteme
Sept. 1998 Vordiplom in Informatik
Mai 2002 Diplom in Informatik

Berufstätigkeit

2002 – 2008 Wissenschaftlicher Angestellter
Lehrstuhl für Informatik 5, RWTH Aachen
2007 – 2008 Lehrauftrag „Datenbanken“
Fachhochschule Aachen
2008 – 2009 Dozent für Informatik
German University of Technology (Muscat, Oman)
seit 2009 Wissenschaftlicher Angestellter
Lehrstuhl für Informatik 5, RWTH Aachen