

Adaptive Verfahren zur nutzerzentrierten Organisation von Musiksammlungen

Sebastian Stober

Data & Knowledge Engineering Gruppe

Fakultät für Informatik, Otto-von-Guericke-Universität, Magdeburg

Sebastian.Stober@ovgu.de

Abstract: Music Information Retrieval (MIR) Systeme müssen fazettenreiche Informationen verarbeiten und gleichzeitig mit heterogenen Nutzern umgehen können. Insbesondere wenn es darum geht, eine Musiksammlung zu organisieren, stellen die verschiedenen Sichtweisen der Nutzer, verursacht durch deren unterschiedliche Kompetenz, musikalischen Hintergrund und Geschmack, eine große Herausforderung dar. Diese Herausforderung wird hier adressiert, indem adaptive Verfahren für verschiedene Elemente von MIR Systemen vorgeschlagen werden: Datenadaptive Techniken zur Merkmalsextraktion werden beschrieben, welche zum Ziel haben, die Qualität und Robustheit der aus Audioaufnahmen extrahierten Informationen zu verbessern. Das klassische Problem der Genreklassifikation wird aus einer neuen nutzerzentrierten Sichtweise behandelt – anknüpfend an die Idee idiosynkratischer Genres, welche die persönlichen Hörgewohnheiten eines Nutzer besser widerspiegeln. Eine adaptive Visualisierungstechnik zur Exploration und Organisation von Musiksammlungen wird entwickelt, die insbesondere Darstellungsfehler adressiert, welche ein weit verbreitetes und unumgängliche Problem von Techniken zur Dimensionsreduktion sind. Darüber hinaus wird urmissen, wie diese Technik eingesetzt werden kann, um die Interessantheit von Musikempfehlungen zu verbessern, und neue blickbasierte Interaktionstechniken ermöglicht. Schließlich wird ein allgemeiner Ansatz für adaptive Musikähnlichkeit vorgestellt, welcher als Kern für eine Vielzahl adaptiver MIR Anwendungen dient. Die Einsatzmöglichkeiten der beschriebenen Verfahren werden an verschiedenen Anwendungsprototypen gezeigt.

Dank immer fortgeschrittenerer Analysemethoden aus dem Music Information Retrieval (MIR) werden zukünftige Generationen von Programmen zur Verwaltung von Musiksammlungen immer besser den Inhalt der Musikstücke verstehen können. Dadurch wird es möglich, Musikstücke inhaltlich miteinander zu vergleichen und beispielsweise ähnliche Stücke zu gruppieren, anstatt einfach nur nach Genre, Künstler und Album zu sortieren. Dabei ergeben sich jedoch auch neue Herausforderungen und Problemstellungen. Einerseits müssen die MIR Systeme auf der Datenseite mit einer Vielzahl verschiedener Facetten von Musik umgehen (wie z.B. Melodie, Harmonie, Rhythmus, Dynamik, Instrumentierung, Text) und mit Merkmalen, welche die Musikstücke auf ganz unterschiedlichen Ebenen beschreiben und sich dabei auf verschiedene Facetten beziehen. Gleichzeitig muss aber auch mit einer starken Varianz unter den Benutzern eines MIR Systems gerechnet werden mit Unterschieden im Hörverhalten, Musikgeschmack und nicht zuletzt im musikalischen Hintergrund. Dies führt beispielsweise dazu, dass im Allgemeinen nicht alle Nutzer eines Systems Musikstücke auf die gleiche (objektive) Art und Weise vergleichen. So mag ein Musiker dazu neigen, stärker auf Strukturen, Instrumentierung oder

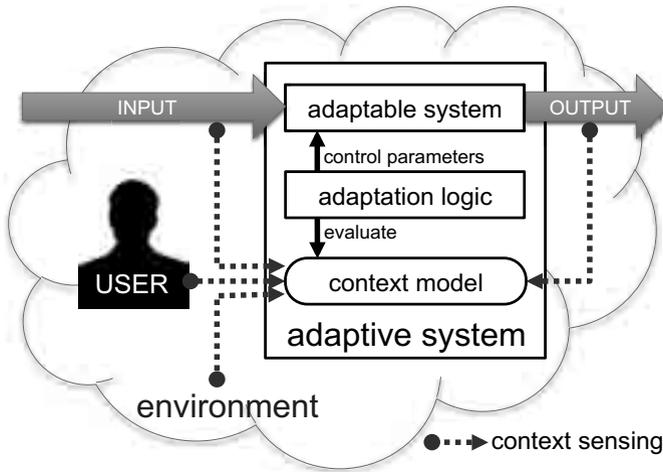


Abbildung 1: Allgemeines Modell eines adaptiven Systems.

Harmonien zu achten und dabei vielleicht seinem Instrument eine besondere Gewichtung geben. Nicht-Musiker werden sich beim Vergleichen möglicherweise eher auf die Klangfarbe oder allgemeine Stimmung eines Stückes stützen. Andere wiederum achten, sofern sie der jeweiligen Sprache mächtig sind, auf den Liedtext. Um mit der großen Diversität in den Musikinformationen und bei den Benutzern von MIR Systemen umzugehen, bietet sich als eine Lösungsmöglichkeit die Verwendung adaptiver Methoden an.

Als Grundlage für die Entwicklung adaptiver MIR-Methoden wurden zunächst die folgenden Definitionen sowie ein allgemeines Modell für adaptive Systeme erarbeitet. Ein System wird als *adaptierbar* bezeichnet, wenn sein Verhalten mittels von außen zugänglicher Parameter manuell angepasst werden kann. Ein *adaptive* System nimmt hingegen diese Anpassung selbständig vor, wobei die Änderung im Verhalten vom Kontext abhängt und zielorientiert sein muss, so dass das Systemverhalten bezüglich eines gegebenen Maßes optimiert wird. Der Kontext ist dabei im weitest möglichen Sinn zu verstehen und umfasst die allgemeine Arbeitsumgebung, den Benutzer und die Ein- und Ausgabedaten (z.B. deren Historie oder statistische Charakteristika). Für die automatische Anpassung muss das System den Kontext (zumindest teilweise) wahrnehmen, ein Kontextmodell als interne Repräsentation konstruieren und schließlich daraus mittels einer Adaptionslogik geeignete Systemänderungen ableiten. Der Kern des Systems kann dabei wiederum als einfaches adaptierbares System betrachtet werden wie in Abbildung 1 dargestellt. Diese Sichtweise ermöglichte eine systematische Analyse und den Vergleich verschiedenster MIR Verfahren mit Blick auf deren Adaptivität [SN12], welche in Umfang und Systematik bisher einmalig ist. Dieser Überblick weißt auch auf besonders vielversprechende Anwendungsfelder für adaptive Techniken im MIR hin, welche in der Arbeit behandelt wurden und in den folgenden Abschnitten zusammengefasst sind.

1 Adaptive Merkmalsextraktion aus Musikaufnahmen

Bei der Analyse von Musikaufnahmen, dem ersten Schritt des generellen Retrievalprozess, kann mit Hilfe adaptiver Methoden eine höhere Qualität und Robustheit der extrahierten Merkmale erreicht werden. Hier wurden zwei Ansätze entwickelt, die im Rahmen von Diplomarbeiten betreuter Studenten umgesetzt wurden. Der erste Ansatz beschäftigt sich mit dem Problem, wie die Melodie aus einer Stereo-Musikaufnahme extrahiert werden kann [DNS07]. Die Idee dabei ist, zunächst mittels eines weit verbreiteten Karaoke-Filters ein Signal für die Hintergrundmusik zu extrahieren. Dieses Signal wird dann zum Einstellen eines Störfilters verwendet, der versucht, möglichst viel der Hintergrundmusik aus der Aufnahme zu entfernen, während die Melodie erhalten bleiben sollte. Für einen solchen Anwendungsfall ist der Störfilter jedoch an sich nicht vorgesehen, da ein weitestgehend konstantes Störsignal vorausgesetzt wird, während die Hintergrundmusik hingegen oft starke zeitliche Veränderlichkeit aufweist. Abhilfe schafft hier, den Störfilter als adaptierbaren Kern in einem adaptiven System einzubetten, welches die Filterparameter dynamisch bezüglich eines kleinen Zeitfensters wählt. Somit wird zu jedem Zeitpunkt nur ein kleiner zeitlich begrenzter Ausschnitt der Hintergrundmusik zur Anpassung des Filters verwendet, was der Annahme eines konstanten Störsignals deutlich näher kommt und schließlich zu einem besseren Filterergebnis führt.

Der zweite Ansatz beschäftigt sich mit der Fehlerkorrektur bei der Erkennung von Akkorden in Musikaufnahmen [RSN08]. Die harmonische Akkordfolge ist ein wichtiges Merkmal zur Indexierung und Analyse westlicher Musik. Daher beschäftigt sich eine Vielzahl von Arbeiten im MIR mit dieser Problematik. Trotz großer Fortschritte sind Klassifikationsfehler jedoch keine Seltenheit. Dies liegt unter anderem an der Verwendung von Musikinstrumenten und Effekten, die nicht nur harmonische Signalanteile erzeugen. Basierend auf einer umfangreichen Studie existierender Ansätze zur Akkorderkennung, wurde hier eine Unterteilung in drei Phasen vorgenommen: Merkmalsextraktion, Akkordklassifikation und Nachverarbeitung. Anschließend wurde ein adaptives Verfahren für die Nachverarbeitung vorgeschlagen, welches unabhängig von den ersten zwei Arbeitsschritten angewendet werden kann. Dabei werden (ungenau) Informationen über die Akkorde in der (zeitlichen) Nachbarschaft als Kontext verwendet, um mögliche Fehlklassifikationen zu korrigieren. Dazu bildet ein probabilistischer Klassifikator den adaptierbaren Kern des adaptiven Systems. Als Systemparameter werden die Marginalwahrscheinlichkeiten der einzelnen Akkorde dynamisch bezüglich der Häufigkeit in der Nachbarschaft angepasst und so der Klassifikationsprozess beeinflusst. Dies führte zu einer deutlichen Verbesserung der Klassifikationsgenauigkeit in Experimenten mit drei verschiedenen Basis-Klassifikationsverfahren.

2 Genreklassifikation mit nutzerspezifischen Genrekategorien

Die Einordnung von Musikstücken in Genres ist ein häufig verwendeter Ansatz zur Sortierung von Musiksammlungen – insbesondere für große Kataloge im Handel. Eine solche Sortierung ist jedoch oftmals nur begrenzt hilfreich, da einerseits generische Genres wie

“Rock” oder “Pop” zu undifferenziert sind und es andererseits aber auch schwierig ist, für sehr spezifische Genres wie “Scottish Lo-Fi Post-Rock” einen Konsens unter verschiedenen Nutzern zu finden. Es wurde daher ein alternativer Ansatz untersucht: Anstatt Musik in mehr oder weniger künstliche Schubladen zu pressen, könnte ein nutzeradaptives MIR System Genrekategorien lernen, die auf den individuellen Nutzer zugeschnitten und somit sinnvoll und intuitiv verständlich sind. Verschiedene frühe Studien aus dem MIR deuten zumindest darauf hin, dass sich sinnvolle nutzerbezogene Genrekategorien aus dem individuellen Nutzungsverhalten ableiten lassen wie z.B. “Musik zum Autofahren”. In einer Vorstudie im Rahmen der Dissertation wurde daher zunächst ein Prototyp zur Aufzeichnung von einfachen Wetterinformationen als Hörkontext entworfen und mit einer kleinen Nutzergruppe getestet. Außerdem wurde eine Reihe von weitergehenden Möglichkeiten zur automatischen Aufzeichnung von Kontextinformationen vorgeschlagen, die mit einfachen technischen Mitteln realisierbar wären. Da jedoch einige dieser Möglichkeiten stark in die Privatsphäre der Nutzer eingreifen, wurde zunächst eine umfangreiche Studie zur Akzeptanz der Aufzeichnungstechniken durchgeführt [SSN09]. Im Rahmen der CeBIT 2009 und mit Hilfe eines Online-Fragebogens wurden insgesamt 461 Personen befragt. Die Umfrageergebnisse zeigen deutlich, dass die potentiellen Nutzer der Aufzeichnung von Hörkontextinformation sehr kritisch gegenüber stehen und ihre Privatsphäre Vorrang hat. Als generelle Richtlinie für die zukünftige Entwicklung personalisierter MIR Anwendungen kann aus den Antworten außerdem gefolgert werden, dass die Benutzer jederzeit die volle Kontrolle haben müssen – sowohl über die aufgezeichneten Kontextinformationen als auch darüber, ob und inwieweit diese zur Adaption verwendet werden.

3 Fokusadaptive Visualisierung von Musiksammlungen

Die Visualisierung von Musiksammlungen stellt den ersten der beiden Hauptschwerpunkte dieser Arbeit dar. Viele Ansätze zur Visualisierung einer Musiksammlung basieren auf Techniken, bei denen Objekte (Musikstücke, Alben oder Künstler) aus einem hochdimensionalen Merkmalsraum für die Darstellung in den zwei- oder dreidimensionalen Raum abgebildet werden. Ziel dabei ist es, die Objekte in einer Art Karte der Sammlung so anzuordnen, dass benachbarte Objekte einander sehr ähnlich sind und die Ähnlichkeit mit wachsendem Abstand in der Karte abnimmt. Dabei kommt es durch die Dimensionsreduktion zwangsläufig zu Verzerrungen der Abstände. Als Folge kann es vorkommen, dass benachbarte Objekte sich gar nicht so sehr ähneln, wie es die Darstellung in der Karte vermuten lässt, oder weit von einander entfernte Objekte sehr ähnlich sind. Im letzteren Fall sind beide Regionen durch eine Art “Wurmloch” (über den hochdimensionalen Merkmalsraum) miteinander verbunden und gehören ursprünglich im Merkmalsraum zur gleichen Nachbarschaft. Im Rahmen der Arbeit wurde daher die fokusadaptive SpringLens entwickelt, eine interaktive Visualisierungstechnik, die eine globale Sicht auf eine Musiksammlung ermöglicht und mit adaptiven Filterfunktionen und multifokalem Zoom die beschriebenen Verzerrungsprobleme gezielt adressiert. (Der Name der Technik leitet sich aus dem ihr zugrundeliegenden Verfahren zur nichtlinearen Verzerrung von Bildern [GGSS06] ab.) Dabei wird speziell auf das Phänomen der Wurmlöcher eingegangen, welches bei der

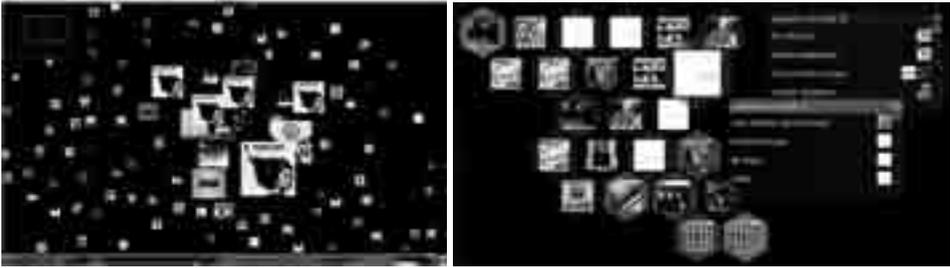


Abbildung 2: Links: MusicGalaxy Prototyp zur Exploration von Musiksammlungen mit Hilfe der fokusadaptiven SpringLens Visualisierung. Rechts: BeatlesExplorer Prototyp zur interaktiven Strukturierung der Musik der Beatles mit einer wachsenden selbstorganisierenden Karte. Demo-Videos zu den Prototypen sind unter <http://www.dke-research.de/aucoma/thesis> verfügbar.

Exploration besonders problematisch ist, da Nutzer im Falle verzerrter Nachbarschaften schnell relevante Objekte übersehen können. Zur Visualisierung der Wurmlöcher in der Kartendarstellung wird eine adaptive multifokale Fischaugenlinse verwendet. Diese besteht aus einem Primärfokus, welcher vom Nutzer gesteuert wird, und einem datengetriebenen Sekundärfokus. Beim Primärfokus handelt es sich um eine gewöhnliche Fischaugenlinse. Mit dieser kann der Nutzer in eine Region hineinzoomen, die ihn interessiert. Die Verzerrung der Linse führt dazu, dass mehr Platz für die Darstellung von Details in dieser Region geschaffen wird, indem die anderen (weniger interessanten) Regionen nach außen gedrückt und komprimiert dargestellt werden. So kann sich der Nutzer die Region von Interesse genauer anschauen, ohne deren Einordnung in den Kontext der gesamten Sammlung zu verlieren. Der Sekundärfokus umfasst mehrere solcher Fischaugenlinsen, die jedoch kleiner sind und nicht direkt vom Nutzer sondern vom System dynamisch in Abhängigkeit vom Primärfokus angepasst werden: Ändert sich der Primärfokus, wird im Hintergrund eine Nächste-Nachbar-Suche im ursprünglichen Merkmalsraum initiiert. Werden nächste Nachbarn zu den Objekten im Primärfokus zurückgeliefert, die sich nicht in direkter Nachbarschaft befinden, werden an den entsprechenden Stellen Sekundärlinsen eingefügt. In der resultierenden verzerrten Darstellung rücken dadurch die entfernten nächsten Nachbarn mit in den Fokus und näher an die Region von Interesse. Basierend auf dieser fokusadaptiven Visualisierungstechnik wurde der Prototyp “MusicGalaxy” [SN10] entwickelt, welcher in Abbildung 2 (links) gezeigt wird. Dabei werden die Musikstücke einer Sammlung als Sterne einer Galaxie visualisiert. Das zugrundeliegende Ähnlichkeitsmaß für die Abstandsberechnung im Merkmalsraum unterstützt eine flexible Anzahl von Facetten, deren Gewichtung adaptierbar ist. Dank spezieller Datenstrukturen können bei einer Änderung des Maßes die Positionen der Musikstücke in der Galaxiekarte in Echtzeit aktualisiert werden wodurch eine interaktive Exploration ermöglicht wird. Zentraler Aspekt ist hier jedoch zunächst die fokusadaptive Visualisierung während sich der folgende Abschnitt dann im Detail mit den Ähnlichkeitsmaßen beschäftigt. Die Entwicklung von MusicGalaxy erfolgte in einem nutzerzentrierten Designprozess. Auf der CeBIT 2010 wurden Meinungen von 112 Besuchern zum ersten interaktiven Prototyp gesammelt. Die nächste überarbeitete Version wurde mit drei Personen ausgiebig getestet

und im Anschluss weiter verbessert. Abschließend wurde eine vergleichende Studie mit Hilfe eines Eyetrackers durchgeführt, an der 30 Personen teilnahmen. Die Ergebnisse der Studie belegen die Nützlichkeit des Sekundärfokus zur Exploration. Im Vergleich mit der herkömmlichen “Pan & Zoom” Technik, die als Standard für kartenbasierte Nutzerschnittstellen betrachtet werden kann, wurden von der fokusadaptiven SpringLens durchweg bessere Werte für Nützlichkeit, Benutzbarkeit und Intuitivität erreicht. Weiterhin konnten mit Hilfe des Eyetrackers verschiedene Navigationsstrategien der Benutzer identifiziert werden. Wie ein weiterer Prototyp zur Exploration von Fotosammlungen beweist, beschränkt sich der Einsatzbereich der entwickelten Visualisierungstechnik nicht nur auf Musikdaten. Verschiedene weiterführende Nutzungsszenarien werden in Abschnitt 5 diskutiert.

4 Kontextadaptive Musikähnlichkeit

Musikähnlichkeit ist der Schlüssel für eine Vielzahl von MIR Anwendungen und stellt daher den zweiten Schwerpunkt dieser Arbeit dar. Ähnlichkeitsmaße werden beispielsweise benötigt, um die Ergebnisliste einer Suche zu sortieren, um bei der Organisation ähnliche Musikstücke zu gruppieren oder um ähnliche Stücke empfehlen zu können. Wie eingangs motiviert, hat Musik jedoch viele Facetten, die nicht zwangsläufig für alle Nutzer gleich wichtig sein müssen. Vielmehr kann die Wichtigkeit einzelner Facetten beispielsweise in Abhängigkeit vom musikalischen Hintergrund des Nutzers oder dessen Retrievalaufgabe stehen. Im Rahmen der Arbeit wurde daher zunächst ein Modell vorgeschlagen, welches subjektive Musikähnlichkeit durch ein parametrisierbares Abstandsmaß mit einer beliebigen Anzahl von Facetten umsetzt. Um den *subjektiven* Abstand zweier Musikstücke zu bestimmen, wird dabei pro Facette ein *objektiver* Abstand berechnet und mit einem subjektiven Faktor zwischen 0 und 1 gewichtet zum Gesamtabstand aufaddiert. Durch Wahl geeigneter Facettengewichte ist dieses Abstandsmaß adaptierbar nach den Vorstellungen des Nutzers. Dank des einfachen linearen Modells ist es zudem auch leicht verständlich. Jedoch kann nicht immer davon ausgegangen werden, dass der Nutzer sich der von ihm beim Vergleich zweier Musikstücke vorgenommenen Facettengewichtung auch bewusst ist. Daher wurde auch ein allgemeiner Ansatz entwickelt, der es einem MIR System erlaubt, die Gewichte aus der Interaktion mit dem Nutzer zu lernen [Sto11]. Der Lernprozess wird dabei als Optimierungsproblem (unter Bedingungen) oder alternativ als binäres Klassifikationsproblem modelliert. In beiden Fällen wird das Verfahren durch relative Abstandsbedingungen gesteuert, die als atomare Informationseinheiten das Kontextmodell bilden. Drei Beispielanwendungen veranschaulichen, wie solche relativen Abstandsbedingungen aus verschiedenen realen Interaktionsszenarien abgeleitet werden können. Bei der ersten Anwendung wurden in Zusammenarbeit mit Forschern des Meertens Instituts in Utrecht Abstandsmaße zur Klassifikation von Volksliedern gelernt. Dazu wurden Annotationen der Experten ausgewertet. Bei der zweiten Anwendung, dem “BeatlesExplorer” (Abbildung 2, rechts), handelt es sich um einen weiteren Prototyp, der im Rahmen dieser Dissertation entwickelt wurde. Er dient zur Exploration der Musik der Beatles und berücksichtigt mehr als 20 Facetten wie z.B. Text, Harmonie, Rhythmus, Instrumentierung und Produzenten. Mit Hilfe einer wachsenden selbstorganisierenden Karte (Growing Self-

Organizing Map; GSOM), werden ähnliche Musikstücke in hexagonale Zellen zusammengefasst, wobei benachbarte Zellen wiederum ähnlich zueinander sind. Durch Drag&Drop kann der Nutzer interaktiv nach seinen Vorstellungen Musikstücke in andere Zellen verschoben. Daraus lassen sich relative Abstandsbedingungen ableiten und das Abstandsmaß anpassen. Die dritte Beispielanwendung ist der bereits vorgestellte Prototyp MusicGalaxy. Hier können Nutzer Objekte durch Taggen in Gruppen einordnen, woraus sich ebenfalls relative Abstandsbedingungen ableiten lassen. Zur Adaption der Facettengewichte bezüglich des Kontextmodells werden schließlich verschiedene Verfahren beschrieben mit leicht unterschiedlichen Zielfunktionen vorgeschlagen, die passend zum Anwendungsszenario ausgewählt werden können. In einer abschließenden Evaluierung mit dem MagnaTagATune Datensatz¹ werden die Verfahren miteinander verglichen. Damit wird von der Modellierung bis hin zur Anwendung und Evaluierung ein umfassender Rahmen für die Entwicklung von MIR Systemen basierend auf einem adaptiven Ähnlichkeitsmaß beschrieben. Darüber hinaus können die entwickelten Techniken auch problemlos auf andere Medien wie z.B. Bilder angewendet werden, solange geeignete Abstandsmaße für die entsprechenden Facetten vorhanden sind.

5 Weiterführende Arbeiten

5.1 Bisoziative Exploration von Musiksammlungen

In Zusammenarbeit mit Stefan Haun wurde im Rahmen des EU-Projektes “BISON”² ein Ansatz zur Entdeckung von Bisoziationen in großen Informationsräumen entwickelt [SHN11]. Der Begriff der Bisoziation geht auf den Künstler Arthur Köstler zurück [Kös64] und bezeichnet eine Assoziation, bei der Domaingrenzen überschritten werden und somit eine nicht offensichtliche Verknüpfung hergestellt wird. Durch Zweckentfremdung des Sekundärfokus der fokusadaptiven SpringLens können Bisoziationen zwischen Objekten in einer Sammlung hervorgehoben werden. Am Beispiel von MusicGalaxy wird gezeigt, wie sich dadurch interessante und überraschende Musikempfehlungen finden lassen. Die Technik lässt sich jedoch auch ohne Weiteres auf andere Anwendungsszenarien übertragen. Die grundlegende Idee dabei ist, zwei grundverschiedene Sichten auf die Sammlung in der Visualisierung miteinander zu kombinieren. Die primäre Sicht wird dazu direkt durch die Galaxiekarte dargestellt und kann durch die Anpassung des zugrundeliegenden Abstandsmaßes verändert werden. Die sekundäre Sicht wird hingegen indirekt durch den Sekundärfokus der SpringLens visualisiert, welcher nun nächste Nachbarn bezüglich der sekundären Sicht hervorhebt. Dabei kann die sekundäre Sicht verschiedene Formen haben. Zum einen ist es möglich, ein alternatives Abstandsmaß zu verwenden – beispielsweise eines mit Facettengewichten, die orthogonal zum Abstandsmaß der primären Sicht gewählt sind (Vgl. Abbildung 3). So könnte z.B. ein Abstandsmaß, welches klangliche Eigenschaften der Musikstücke berücksichtigt, für die Berechnung der Karte mit einem sekundären

¹<http://tagatune.org/Magnatagatune.html>

²Bisociation Network for Creative Information Discovery, <http://www.bisonet.eu/>

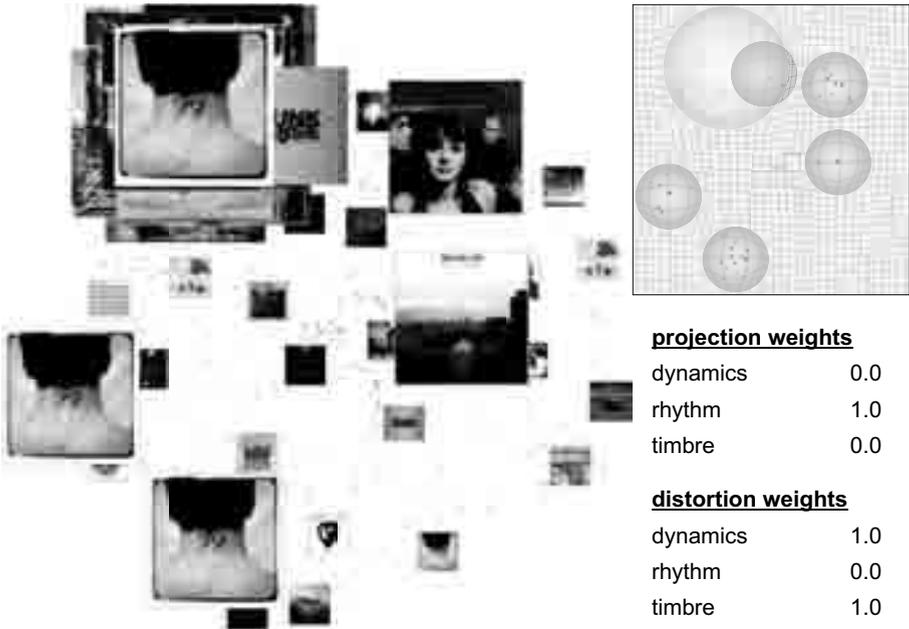


Abbildung 3: Kombination von zwei Abstandsmaßen mit orthogonalen Facettengewichten. Links: MusicGalaxy (invertiertes Farbschema). Rechts oben: Entsprechende Linsenverzerrung bestehend aus vom Benutzer kontrollierten Primärfokus (rot) und adaptiven Sekundärfokus mit mehreren kleinen Linsen (blau). Rechts unten: Facettengewichte zur Berechnung der Galaxiekarte (projection weights) und zur Erkennung nächster Nachbarn für die Verzerrung (distortion weights).

Abstandsmaß, welches die Songtexte vergleicht, kombiniert werden. Lenkt der Benutzer nun den Primärfokus auf ein Musikstück, wird der Sekundärfokus textlich ähnliche Stücke hervorheben, wobei in der Karte weit entfernte Stücke besonders interessant sind, da sie ein ähnliches Thema anders musikalisch umsetzen. Durch Navigation über den Sekundärfokus kann der Benutzer somit schrittweise an sowohl musikalisch als auch textlich ansprechende Musikstücke herangeführt werden. Alternativ ist es auch möglich, für die sekundäre Sicht eine Graphstruktur zu verwenden, in der Beziehungen zwischen Objekten explizit in Form von Kanten vorliegen. Nächste Nachbarn können hier direkt durch Traversieren des Graphen gefunden werden. Diese Konstellation ist besonders interessant, weil sie die Zusammenführung von karten- und graphbasierten Ansätzen ermöglicht, die sonst nur getrennt betrachtet werden, weil sie zwei völlig unterschiedlichen Explorationsstrategien entsprechen. MusicGalaxy konstruiert beispielsweise aus den Informationen der Musicbrainz Datenbank³ einen Graph für die Musikstücke in der Sammlung, der zur Steuerung des Sekundärfokus verwendet werden kann. Dabei werden unter anderem auch Beziehungen zu Künstlern, Alben, Plattenlabeln mit einbezogen.

³<http://musicbrainz.org/>

5.2 Blickgesteuerter Adaptiver Fokus

In Zusammenarbeit mit Sophie Stellmach wurde auf Basis der fokusadaptiven SpringLens schließlich eine blickgestützte Benutzerschnittstelle zur Exploration von Mediensammlungen umgesetzt. Die grundlegende Idee ist hier, dass der Primärfokus statt mit der Maus direkt mit dem Blick gesteuert werden kann. Dabei kommt ein Eyetracker zum Einsatz, der bereits zur Evaluierung der fokusadaptiven SpringLens Visualisierung genutzt wurde. Blickinformationen bieten sich als natürliche Eingabemethode an, da der Blick oft einer manuellen Aktion vorausgeht. Die Verwendung von Fischaugenlinsen in der fokusadaptiven SpringLens hat zudem den Vorteil, dass dadurch Objekte von Interesse lokal vergrößert werden und somit durch den Blick auch leichter ausgewählt werden können. Bei rein blickgesteuerten Benutzerschnittstellen tritt jedoch das sogenannte “Midas Touch” Problem [Jac90] auf, welches nach König Midas aus der griechischen Mythologie benannt ist, der alles, was er anfasste, in Gold verwandelte – ob er wollte oder nicht. Hier bezieht sich dies auf die Schwierigkeit, gewollte von ungewollten Aktionen zu unterscheiden. Um diesem Problem entgegenzuwirken, wurde hier eine zweite Eingabemodalität hinzugenommen. Untersucht wurde dabei die Kombination mit einer Tastatur (repräsentativ für Tasteneingabegeräte wie z.B. Fernbedienungen) und mit einem Smartphone (mit Lagesensor und Touch-Eingabe). Durch eine frühe Einbindung von Benutzern in einen nutzerzentrierten Designprozess konnten entsprechend intuitive und natürliche Interaktionstechniken herausgearbeitet werden. Diese wurden schließlich im “GazeGalaxy” Prototyp umgesetzt, der eine Erweiterung von MusicGalaxy darstellt. Die Ergebnisse einer ersten Nutzerstudie mit diesem Prototyp belegen, dass der Blick tatsächlich als natürlicher Eingabekanal dienen kann und die Verwendung zur Steuerung der Fischaugenlinse als intuitiv betrachtet wird, solange bestimmte grundlegende Designrichtlinien berücksichtigt werden: Erstens sind Blickdaten von Natur aus ungenau und daher sollte die Interaktion nicht von genauen Positionen abhängen. Zweitens sollten Benutzer ihre Aktionen durch zusätzliche explizite Befehle bestätigen können, um ungewollte Aktionen zu vermeiden. Der resultierende Konferenzbeitrag [SSDN11] erhielt auf der NGCA 2011 den Best Paper Award. An weiteren Verbesserungen der Interaktionstechniken wird derzeit gearbeitet.

Danksagung

Die hier zusammengefasste Dissertation fand im Rahmen des DFG-finanzierten Projektes “AUCOMA - Adaptive User-Centered Organisation of Music Archives” statt und wurde zusätzlich durch ein Promotionsstipendium der Studienstiftung des deutschen Volkes gefördert. Die Entwicklung der bisoziativen SpringLens wurde durch die Europäische Kommission unterstützt (FP7-ICT-2007- C FET-Open, contract no. BISON-211898).

Literatur

- [DNS07] Alexander Duda, Andreas Nürnberger und Sebastian Stober. Towards Query by Singing/Humming on Audio Databases. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR'07)*, 2007.
- [GGSS06] Tobias Germer, Timo Götzelmann, Martin Spindler und Thomas Strothotte. SpringLens: Distributed Nonlinear Magnifications. In *Eurographics 2006 - Short Papers*, 2006.
- [Jac90] Robert J. K. Jacob. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*, 1990.
- [Kös64] Arthur Köstler. *The Act of Creation*. Macmillan, 1964.
- [RSN08] Johannes Reinhard, Sebastian Stober und Andreas Nürnberger. Enhancing Chord Classification through Neighbourhood Histograms. In *Proceedings of the 6th International Workshop on Content-Based Multimedia Indexing (CBMI'08)*, 2008.
- [SHN11] Sebastian Stober, Stefan Haun und Andreas Nürnberger. Creating an Environment for Bisociative Music Discovery and Recommendation. In *Proceedings of Audio Mostly 2011 – 6th Conference on Interaction with Sound*, 2011.
- [SN10] Sebastian Stober und Andreas Nürnberger. A Multi-Focus Zoomable Interface for Multi-Facet Exploration of Music Collections. In *Extended Proceedings of 7th International Symposium on Computer Music Modeling and Retrieval (CMMR'10)*, 2010.
- [SN12] Sebastian Stober und Andreas Nürnberger. Adaptive Music Retrieval - A State of the Art. *Multimedia Tools and Applications*, 2012. (Online First Article).
- [SSDN11] Sophie Stellmach, Sebastian Stober, Raimund Dachsel und Andreas Nürnberger. Designing Gaze-supported Multimodal Interactions for the Exploration of Large Image Collections. In *Proceedings of 1st International Conference on Novel Gaze-Controlled Applications (NGCA'11)*, 2011.
- [SSN09] Sebastian Stober, Matthias Steinbrecher und Andreas Nürnberger. A Survey on the Acceptance of Listening Context Logging for MIR Applications. In *Proceedings of the 3rd Workshop on Learning the Semantics of Audio Signals (LSAS'09)*, 2009.
- [Sto11] Sebastian Stober. Adaptive Distance Measures for Exploration and Structuring of Music Collections. In *Proceedings of AES 42nd Conference on Semantic Audio*, 2011.



Sebastian Stober, geboren am 10. Dezember 1980 in Halberstadt, studierte von 2000 bis 2005 Informatik mit Nebenfach Mathematik an der *Otto-von-Guericke-Universität* in Magdeburg mit einem siebenmonatigen Forschungsaufenthalt an der *University of Melbourne* in Australien. Seit 2006 ist er wissenschaftlicher Mitarbeiter in der *Data & Knowledge Engineering Gruppe* mit den Forschungsschwerpunkten Maschinelles Lernen, Musik Information Retrieval und Adaptive Systeme. Er organisierte 2006 den *1st International Workshop on Learning Semantics of Audio Signals (LSAS)* und zwei Folgeveranstaltungen 2007 und 2009. Seit

2007 ist er Mitorganisator des *International Workshop on Adaptive Multimedia Retrieval (AMR)*. Seine Arbeit als Doktorand beendete er 2011 erfolgreich mit seiner Dissertation zum Thema “Adaptive Methods for User-Centered Organization of Music Collections.”