

# SyncPlayer - Multimodale Wiedergabe, Navigation und Suche in heterogenen digitalen Musikkollektionen

David Damm\*, Christian Fremerey\*, Frank Kurth<sup>†</sup>, Meinard Müller<sup>†</sup>, Michael Clausen\*

\*Institut für Informatik III  
Universität Bonn  
D-53117 Bonn

<sup>†</sup>Max-Planck-Institut (MPI)  
für Informatik  
D-66123 Saarbrücken

<sup>‡</sup>Forschungsgesellschaft für  
Angewandte Naturwissenschaften  
D-53343 Wachtberg

## Abstract

Durch systematische Digitalisierung sind in den letzten Jahren umfangreiche Bestände von Musikdokumenten entstanden, die digitale Inhalte unterschiedlichster Ausprägungen und Formate enthalten. Man denke hier beispielsweise an CD-Aufnahmen diverser Interpreten, MIDI-Daten, gescannte Notenblätter oder Gesangstexte. Neben der reinen Erfassung solcher Daten kommt der automatisierten Erschließung und Nutzbarmachung solch inhomogener und komplexer Musikdatenbestände eine stetig wachsende Bedeutung zu. In dieser Arbeit geben wir einen Überblick zum aktuellen Stand des SyncPlayer-Frameworks, das unter Verwendung aktueller Erschließungs- und Indexierungstechniken unterschiedliche Funktionalitäten zur multimodalen Wiedergabe, Navigation und inhaltsbasierten Suche von Musik realisiert. Die Navigationskomponenten des SyncPlayers basieren auf zuvor berechneten Verlinkungsstrukturen und erlauben mittels geeigneter Benutzerschnittstellen die synchrone Wiedergabe und Darstellung von Audioaufnahmen und Partiturstellen sowie die simultane Ansteuerung unterschiedlicher Aufnahmen desselben Musikstücks und musikalischer Wiederholungen einzelner Aufnahmen. Weiterhin werden zwei SyncPlayer-Komponenten zum Musikretrieval vorgestellt, die eine Suche anhand verschiedener musikrelevanter Aspekte (Volltextsuche in Liedtextbeständen, harmoniebasierte Suche in Audiodatenbeständen) ermöglichen. Schließlich beschreiben wir, wie diese unterschiedlichen Funktionalitäten im Kontext des DFG-Leistungszentrums PROBADO in ein konkretes Bibliothekssystem eingebunden werden sollen.

## 1 Einleitung

Durch zunehmende Digitalisierung von Musikdaten aller Art sind in den letzten Jahren umfangreiche, oft unstrukturierte Musikdatenbeständen entstanden. In realen Anwendungsszenarien sind diese Bestände im allgemeinen heterogen und enthalten Bild-, Ton- und Textinhalte unterschiedlicher Formate. Im Kontext digitaler Musikbibliotheken spielen dabei CD-Aufnahmen und gescannte Partituren als typische, automatisiert generierbare Digitalisatormen eine wichtige Rolle. Darüber hinaus gibt es eine

wachsende Anzahl von musikbezogenen Textdokumenten (Gesangstexte, Libretti), automatisiert generierter Annotationsdokumente zur Kodierung repräsentativer Passagen (Thumbnails) oder Zusammenfassungen (Summaries) sowie partiturnaher Musikdokumente in digitalen Formaten wie MIDI [Huber, 1999] oder MusicXML [Good, 2001]. Siehe auch [Selfridge-Field, 1997] für eine Zusammenfassung symbolischer Musikdatenformate. Im Anschluss an die reine Erfassung und Digitalisierung von Musikdaten ist in den meisten Anwendungsszenarien die automatisierte Erschließung und Nutzbarmachung der Daten von höchster Wichtigkeit. Hierzu werden zum einen Methoden zur automatisierten Annotation, Verschlagwortung, Verlinkung und Indexierung der unterschiedlichen Musikdokumente benötigt. Zum anderen sind auf dieser Basis Werkzeuge und Benutzerschnittstellen zur adäquaten Präsentation, zur inhaltlichen Suche und zur Navigation in Musikdokumenten- und -kollektionen unter Berücksichtigung unterschiedlichster Modalitäten zu entwickeln.

Die Forschung auf dem Gebiet des *Music Information Retrieval* (MIR) hat im letzten Jahrzehnt vielfältige Methoden zur automatisierten Musikdatenerschließung und inhaltsbasierten Musiksuche hervorgebracht. Zu den Erschließungsmethoden gehören insbesondere unterschiedliche Synchronisationsverfahren zur automatischen Verlinkung zweier Datenströme unterschiedlicher Formate. Hierzu zählen Alignmentmethoden zur Verlinkung von CD-Aufnahmen mit symbolischen Notenformaten wie MIDI [Arifi *et al.*, 2004; Dixon and Widmer, 2005; Hu *et al.*, 2003; Raphael, 2004; Soulez *et al.*, 2003; Turetsky and Ellis, 2003], Gesangstexten [Wang *et al.*, 2004] oder gescannten Notenbildern [Kurth *et al.*, 2007]. Weiterhin wurden unterschiedliche Algorithmen zur Bestimmung der Wiederholungsstruktur und repräsentativer Ausschnitte entwickelt, siehe u. a. [Bartsch and Wakefield, 2005; Goto, 2003; Maddage *et al.*, 2004; Müller and Kurth, 2007; Peeters *et al.*, 2002]. Neben traditionellen textbasierten Verfahren zur Musiksuche, die auf Annotationen und symbolischen Musikdaten [Clausen and Kurth, 2004; Typke *et al.*, 2005] basieren, wurden in den letzten Jahren auch inhaltsbasierte Suchverfahren entwickelt, die unmittelbar auf den Audiodaten arbeiten, siehe u. a. [Cano *et al.*, 2002; Kurth and Müller, 2008]. Für einen Überblick zu Fragestellungen der automatisierten Musikdatenerschließung verweisen wir auch auf [Klapuri and Davy, 2006; Müller, 2007; Pardo, 2006].

Demgegenüber fehlen bislang geeignete Content Management Systeme für digitale Musikkollektionen, die die verfügbaren Erschließungstechniken mit geeigneten Werkzeugen und Benutzerschnittstellen zur multimodalen Wiedergabe, Navigation und Suche systematisch verbinden. In

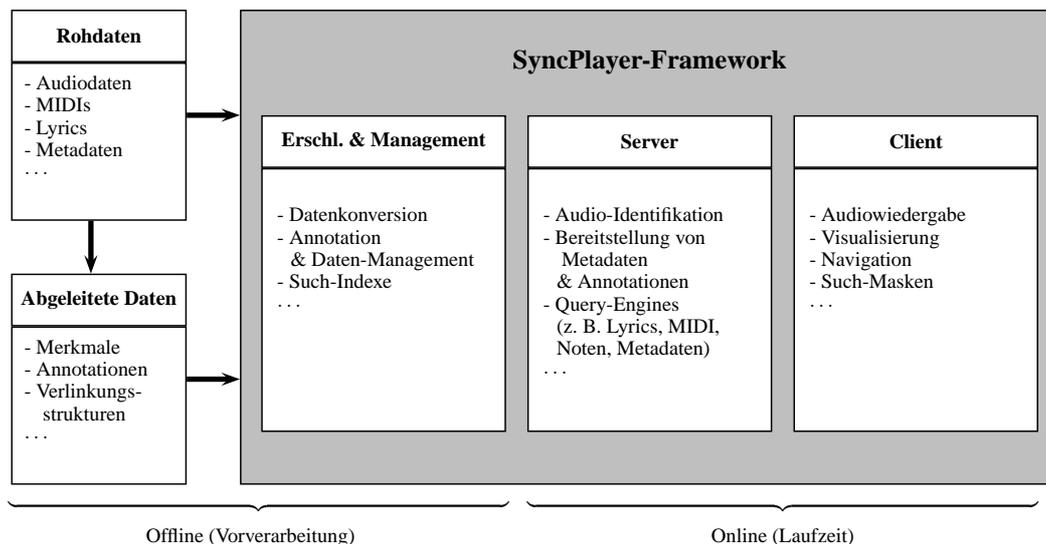


Abbildung 1: Gesamtübersicht über das SyncPlayer-Framework.

dieser Arbeit geben wir einen Überblick über den aktuellen Stand des SyncPlayer-Frameworks, das unter Verwendung aktueller MIR-Technologien Teile dieser Lücke zu schließen versucht. Das erstmals in [Kurth *et al.*, 2005] vorgestellte Framework, welches in den letzten Jahren um zahlreiche Funktionalitäten und Schnittstellen erweitert wurde, erlaubt dabei die systematische Integration von durch Erschließungswerkzeuge abgeleiteten Daten — angefangen von Verlinkungs- und Strukturdaten bis hin zu automatisiert erstellten Suchindizes. Mittels geeigneter Nutzerschnittstellen gestattet das System dann eine multimodale sowie effiziente Präsentation, Navigation und Suche in den Dokumenten der vorliegenden Musikkollektion.

Während Abschnitt 2 den technologischen Hintergrund des SyncPlayer-Frameworks zusammenfasst, diskutieren die darauffolgenden Abschnitte jeweils einzelne Systemkomponenten und spezifische Benutzerschnittstellen. In Abschnitt 3 geht es um eine Anwendung, die auf einer Verlinkung von CD-Aufnahmen und gescannten Partiturdaten basiert [Kurth *et al.*, 2007]. Die Verlinkungsdaten ermöglichen zum einen eine zeitsynchrone Darstellung von Partiturstellen während der Wiedergabe einer CD-Aufnahme und zum anderen eine multimodale Navigation in Audio- und Partiturdaten. Abschnitt 4 diskutiert eine Anwendung, bei der während der Audiowiedergabe zwischen unterschiedlichen Audioaufnahmen desselben Musikstücks nahtlos übergeblendet werden kann (*Inter-Dokument-Navigation*). In Abschnitt 5 wird dann eine Benutzerschnittstelle beschrieben, die die Navigation innerhalb einer einzelnen Audioaufnahme ermöglicht (*Intra-Dokument-Navigation*). Anhand einer automatisch erstellten und durch ein Blockdiagramm dargestellten Wiederholungsstruktur [Müller and Kurth, 2007] kann hierbei der Benutzer durch musikalisch ähnliche/identische Audioausschnitte effizient navigieren. Zur inhaltsbasierten Suche stellt die aktuelle Version des SyncPlayers zwei Methoden zur Verfügung. Beim Audiomatching (Abschnitt 6) erhält der Benutzer bei Auswahl eines kurzen Audioausschnitts (in der Regel 10-30 Sekunden) eine Liste aller musikalisch ähnlichen Ausschnitte in sämtlichen Stücken der Musikkollektion [Kurth and Müller, 2008]. Die Lyrics-Suche (Abschnitt 7) erlaubt schließlich eine klassische Volltextsuche anhand von Gesangstexten, wobei die textu-

ellen Trefferstellen zeitlich genau mit zugehörigen Audioaufnahmen verlinkt sind und innerhalb einer multimodalen, karaoke-ähnlichen Benutzerschnittstelle dargestellt werden können [Damm, 2007; Müller *et al.*, 2007].

Das SyncPlayer-Framework ist der Ausgangspunkt und verbindendes Element von zwei aktuellen Forschungsrichtungen. Einerseits liefert unser Framework die Basis zur Erschließung und Nutzbarmachung eines Musik-Repositoriums, das zur Zeit an der Bayerischen Staatsbibliothek (BSB) in München im Kontext des DFG-Leistungszentrums PROBADO [Krottmaier *et al.*, 2007] erstellt wird. Insbesondere stehen hier gescannte Partiturbände und CD-Aufnahmen (als Dokumententypen) sowie die Automatisierung der inhaltlichen Erschließung innerhalb des bibliothekarischen Geschäftsgangs [Diet and Kurth, 2007] im Vordergrund. Andererseits zielt das aktuelle DFG-geförderte ARMADA-Vorhaben durch Intensivierung der Grundlagenforschung auf die Entwicklung robuster, für reale Dokumentenbestände tauglicher Technologien zur Erschließung von Musikdokumenten ab, die schließlich in einem integrierten Content Management System für Musikkollektionen realisiert werden sollen.

## 2 Überblick über das SyncPlayer-Framework

In diesem Abschnitt skizzieren wir zunächst die Basisfunktionalität des SyncPlayer-Frameworks aus Benutzersicht und beschreiben dann die drei Hauptkomponenten des Frameworks im Überblick. Dabei spielt insbesondere das Plug-In-Konzept eine Rolle, innerhalb dessen die in den weiteren Abschnitten beschriebenen Module realisiert sind.

Eine grundlegende Funktionalität des SyncPlayers ist die multimodale, synchronisierte Wiedergabe verschiedener Musikdatenströme. Der SyncPlayer ist hierzu als Client/Server-basierter Audioplayer realisiert, der neben der akustischen Wiedergabe von Audioaufnahmen in der Lage ist, zusätzliche, zu der aktuell wiedergegebenen Audioaufnahme assoziierte Daten vom Server zu beschaffen und dem Benutzer auf geeignete Art und Weise zu präsentieren. Insbesondere handelt es sich hierbei um Daten zur visuellen Präsentation wie gescannte Notenblätter, Gesangstexte, Partiturdaten im MIDI-Format, Akkorde etc.

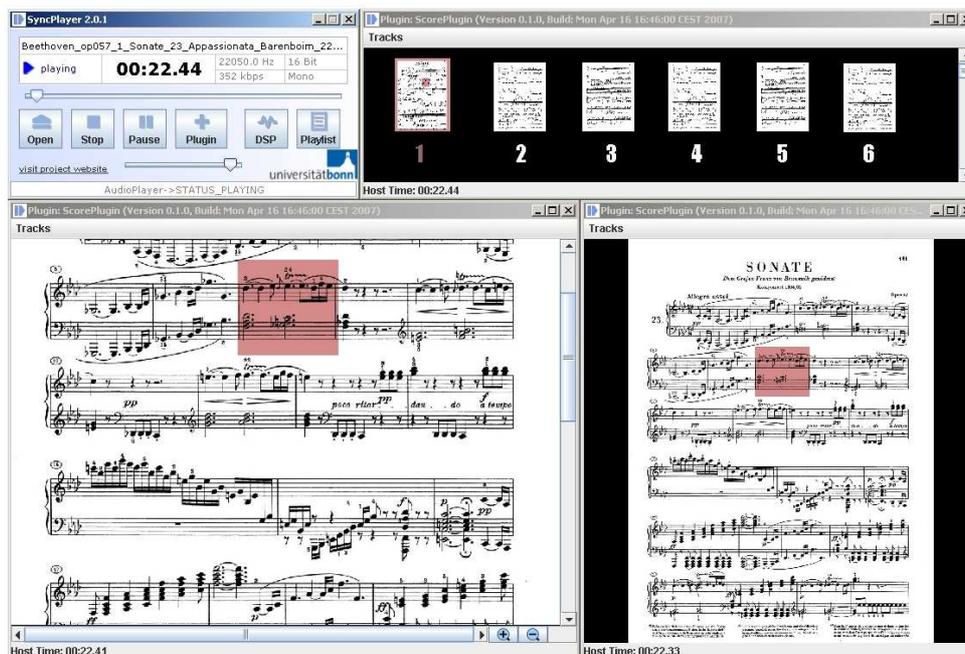


Abbildung 2: SyncPlayer mit mehreren Instanzen des ScoreViewer-Plug-Ins zur Anzeige von Partiturinformatoren auf unterschiedlichen Detailstufen. Der ScoreViewer markiert den in der Audioaufnahme aktuell wiedergegebenen Takt in der zugehörigen Partitur. Weiterhin erlaubt der ScoreViewer eine synchronisierte Navigation innerhalb der Audioaufnahme und zugehörigen Partitur.

Diese werden dem Benutzer zeitsynchron zur akustischen Wiedergabe angezeigt, wodurch unterschiedliche Modalitäten gleichzeitig angesprochen werden.

Das SyncPlayer-Framework lässt sich grob in drei Hauptkomponenten unterteilen, die Client-, Server- sowie Erschließungs- und Management-Komponente. Die Komponenten sind modular aufgebaut und größtenteils in Java implementiert; nur einige zeitkritische Teilkomponenten sind als C++-Module realisiert. Die Funktionalitäten der Hauptkomponenten lassen sich wie folgt zusammenfassend beschreiben (vgl. hierzu auch Abb. 1):

- Die Benutzeroberfläche bildet die *Client-Komponente*, die als Grundfunktionalität einen Audioplayer zur Wiedergabe von MP3- und WAV-Dateien bereitstellt. Erweiterte Funktionalitäten können über eine Plug-In-Schnittstelle angebunden werden. Zu diesen Funktionalitäten zählen Benutzerschnittstellen zur Visualisierung, Interaktion und inhaltsbasierter Suche, die als einzelne, voneinander unabhängige, aber miteinander kombinierbare, Plug-Ins realisiert sind (vgl. Abb. 7).
- Die zu einer wiedergegebenen Audioaufnahme assoziierten Daten, die der Client abrufen kann, werden an zentraler Stelle, der *Server-Komponente*, bereitgestellt. Zu diesen Daten gehören Annotationen, Verlinkungsstrukturen verschiedener Datenströme oder Ergebnisse einer inhaltsbasierten Suche. Die Server-Komponente ist ebenfalls modular aufgebaut, wobei sämtliche Interaktionen der Clients mit dem Server über sogenannte *Query-Engines* abgewickelt werden. Für jede Interaktionsart (zum Beispiel Abfrage von Annotationen, Bearbeitung einer Anfrage) ist dabei eine eigene Query-Engine zuständig.
- Die *Erschließungs- und Management-Komponente*

beinhaltet die verfügbaren Werkzeuge zur Verwaltung von Datenbanken, Indexen und Annotationen. Weiterhin existieren Werkzeuge zur Erstellung von Verlinkungsstrukturen zwischen diversen Musikdokumenten, zur Generierung von Annotationen und zum Aufbau von Suchindexen.

Server-seitig liegt eine große Kollektion von Musikdokumenten unterschiedlicher Formate vor. Im folgenden werden diese auch als *Rohdaten* bezeichnet. Im vorliegenden Szenario bestehen diese überwiegend aus Audioaufnahmen, gescannten Partituren und MIDI-Dokumenten. Aus diesen Rohdaten werden sogenannte *abgeleitete Daten* gewonnen. Diese bestehen aus Audiomerkmale, diversen Synchronisations- und Verlinkungsstrukturen (zum Beispiel Ergebnisse einer Audio-Partitur- oder Audio-Audio-Synchronisation) sowie automatisch extrahierten Wiederholungsstrukturen. Die Verlinkungsstrukturen drücken existierende Relationen zwischen den am Synchronisationsprozess beteiligten Rohdaten aus. Die Wiederholungsstrukturen enthalten zu einzelnen Audioaufnahmen die sich wiederholenden musikalischen Teile.

Mittels der Erschließungs- und Management-Komponente werden die Roh- und abgeleiteten Daten indexiert und in Datenbanken abgelegt, auf die der Server Zugriff hat. Die Erzeugung der abgeleiteten Daten erfolgt in einem Vorverarbeitungsschritt, während der Abruf dieser Daten vom Server zur Laufzeit erfolgt (siehe Abb. 1).

Der Client kann schließlich während der Audiowiedergabe in Echtzeit weitere, inhaltsbezogene Daten sukzessive vom Server abfragen und präsentieren. Hierzu überträgt der Server eine Liste der für eine selektierte Audioaufnahme verfügbaren, in einzelnen *Tracks* organisierten Daten an den Client. Der Benutzer erhält auf Grundlage dieser Trackliste die Möglichkeit, verschiedene Plug-Ins zur Vi-

sualisierung auszuwählen. Derzeit existieren Plug-Ins zur Darstellung von Partiturdaten (in Form von gescannten Notenblättern und im MIDI-Format), Gesangstexten, Akkorden, Audiostrukturdaten sowie Plug-Ins zur Wellenform- und Spektraldarstellung. Darüber hinaus existieren Plug-Ins zur Navigation innerhalb verschiedener Interpretationen eines Musikstücks sowie zur inhaltsbasierten Suche sowohl auf den Audiodaten als auch auf Gesangstexten. Für weitere technische Details, insbesondere die Verwaltung und Implementierung des SyncPlayer-Frameworks betreffend, sei auf [Damm, 2007; Fremerey, 2006; Kurth *et al.*, 2005] verwiesen.

In den folgenden Abschnitten wird eine detailliertere Beschreibung der einzelnen Plug-Ins gegeben, die derzeit im SyncPlayer verfügbar sind.

### 3 Komponente I: ScoreViewer

Der ScoreViewer ist ein Plug-In zur simultanen Anzeige der Partitur eines Musikstücks während der Wiedergabe einer zugehörigen Audioaufnahme. Dem Benutzer wird hier nicht nur das Notenblatt angezeigt, das zum aktuell wiedergegebenen Inhalt der Audioaufnahme korrespondiert, sondern es wird auch der aktuell gespielte Takt hervorgehoben dargestellt (vgl. Abb. 2, links unten). Dadurch kann der Benutzer die Partitur des Musikstücks taktweise mitverfolgen, während er der Audioaufnahme zuhört. Andererseits kann durch die Tatsache, dass bestimmte Zeitabschnitte innerhalb der Audioaufnahme einzelnen Takten innerhalb der Partitur entsprechen, eine bequeme und intuitive Art der Navigation im Musikstück realisiert werden. Durch gezielte Auswahl eines Taktes innerhalb der Partitur wird die Wiedergabe der Audioaufnahme an der zu diesem Takt korrespondierenden Stelle fortgesetzt. Somit hat der Benutzer die Möglichkeit, über die Partitur an beliebige Stellen des Musikstücks zu springen. Man beachte, dass das klassische Vorgehen, eine bestimmte musikalische Stelle innerhalb einer Audioaufnahme manuell zu suchen (zum Beispiel durch Hin- und Herspulen), mit erheblich mehr Zeitaufwand verbunden ist. Bei der hier realisierten Art der Suche entfällt solch ein manuelles Durchlaufen der Audioaufnahme gänzlich.

Abb. 2 zeigt den SyncPlayer (links oben) mit drei Instanzen des ScoreViewer-Plug-Ins. Rechts oben befindet sich der Notenblatt-Browser zur seitenbasierten Navigation innerhalb einer aktuell selektierten Partitur. Wählt der Benutzer hier eine Seite aus, so wird die Wiedergabe der Audioaufnahme entsprechend dem Beginn der Seite gestartet beziehungsweise dort fortgesetzt. Der Benutzer hat weiterhin die Möglichkeit, sich das Notenblatt auf verschiedenen Detailstufen anzeigen zu lassen. So ist rechts unten eine Seitenübersicht über das aktuelle Notenblatt sowie links unten die den aktuell abgespielten Takt umgebende Region des Notenblatts im Ausschnitt dargestellt.

### 4 Komponente II: AudioSwitcher

Das AudioSwitcher-Plug-In bietet die Möglichkeit, während der Audiowiedergabe ad-hoc zwischen verschiedenen Interpretationen eines Musikstücks zu wechseln. Hierbei wird die aktuelle Wiedergabeposition im musikalischen Sinne beibehalten. Genauer bedeutet dies, dass während der Wiedergabe einer Interpretation zu einem beliebigen Zeitpunkt eine andere Interpretation ausgewählt werden kann und die Wiedergabe mit der neu ausgewählten Interpretation an derjenigen Stelle



Abbildung 3: SyncPlayer mit AudioSwitcher-Plug-In. Der AudioSwitcher erlaubt die Auswahl mehrerer Interpretationen eines Musikstücks. In diesem Beispiel kann der Benutzer über die jeweiligen Slider zeitgleich in fünf unterschiedliche Aufnahmen von Beethovens fünfter Symphonie navigieren. Durch Klick auf die zugehörigen Play-Buttons (links) kann zwischen diesen nahtlos hin- und hergewechselt werden.

fortgesetzt wird, die musikalisch äquivalent zu der Stelle innerhalb der alten Interpretation ist, die vor dem Wechsel wiedergegeben wurde. Hierdurch ist beispielsweise ein unmittelbarer Vergleich verschiedener Interpretationen eines Musikstücks möglich, da zwischen den Audioaufnahmen nahtlos und latenzfrei übergeblendet wird (*Inter-Dokument-Navigation*).

Abb. 3 zeigt als Beispiel fünf verschiedene Interpretationen von Beethovens Fünfter Symphonie. Zu jeder Interpretation gehört ein Slider (Zeitleiste) und ein Abspielsymbol (jeweils links neben einem Slider positioniert). Im Beispiel ist als aktuell wiedergegebene Aufnahme die Interpretation von Bernstein (durch ein rotes Abspielsymbol gekennzeichnet) selektiert. Möchte der Benutzer während der Wiedergabe eine andere Interpretation hören, muss er dazu lediglich das jeweilige Abspielsymbol zu einer Interpretation anklicken. Ferner kann der Benutzer durch Ziehen eines Sliders zu der entsprechenden Zeitposition der zugehörigen Interpretation wechseln.

### 5 Komponente III: AudioStructure

Musikstücke aus dem westlichen Kulturkreis lassen sich im allgemeinen in eine Folge musikalisch bedeutsamer Abschnitte unterteilen. Zu solchen Abschnitten gehören bei der populären Musik Intros, Strophen, Refrains und Übergänge. Bei der klassischen Musik sind die möglichen Abschnittsarten und deren weitere Unterstrukturierungen noch vielgestaltiger. Im allgemeinen spricht man aus musikwissenschaftlicher Sicht von der *musikalischen Form* eines Musikstücks.

Obwohl eine automatisierte Analyse der musikalischen Form einer gegebenen Audioaufnahme nach dem derzeiti-

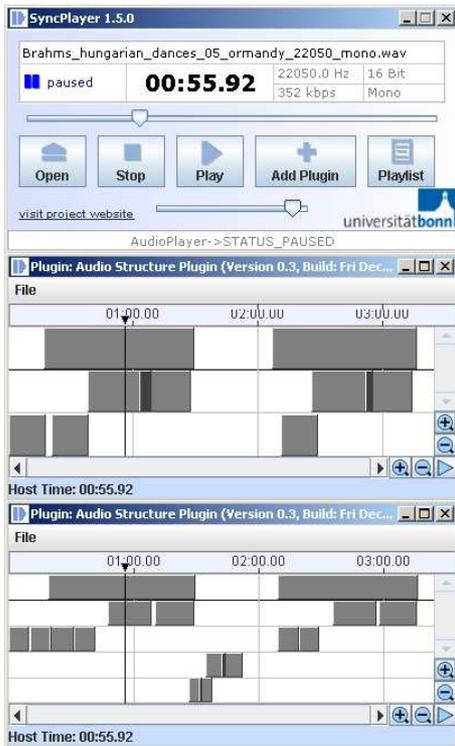


Abbildung 4: SyncPlayer mit zwei Instanzen des AudioStructure-Plug-Ins zur Anzeige von Strukturinformationen. In diesem Beispiel wird die Wiederholungsstruktur einer Ormandy-Aufnahme von Brahms' Ungarischen Tanz Nr. 5 angezeigt, wobei die Blöcke jeder Zeile sich wiederholenden Audiosegmenten entsprechen. Die untere der beiden Instanzen zeigt die Wiederholungsstruktur auf einer feineren Detailstufe.

gen Stand der Technik noch nicht zu bewerkstelligen ist, gibt es einige Ansätze, zumindest eine approximative Musikstruktur einer gegebenen Aufnahme zu berechnen. In einem auf der Analyse der harmonischen Progression basierenden Ansatz [Müller and Kurth, 2007] wird hierzu eine Methode vorgestellt, die zu einer Audioaufnahme versucht, eine vollständige Liste aller sich wiederholenden Abschnitte zu bestimmen. Aufgabe des AudioStructure-Plug-Ins ist es, die so ermittelte Audiostruktur visuell zugänglich zu machen. Dabei dient das Plug-In nicht nur der Anzeige der Struktur, sondern auch einer strukturbasierten Navigation. Ein geeigneter Navigationsmechanismus soll dem Benutzer erlauben, bestimmte Abschnitte direkt ansteuern zu können.

Abb. 4 zeigt den SyncPlayer (oben) mit zwei Instanzen des Plug-Ins (mitte, unten). Die Struktur einer aktuell im SyncPlayer geladenen Audioaufnahme wird in den Plug-Ins durch einzelne Blöcke visualisiert, wobei die Blöcke einer jeden Zeile für einen bestimmten, sich aus Sicht der harmonischen Progression wiederholenden Abschnitt stehen. Das Beispiel in Abb. 4 zeigt die Struktur einer Ormandy-Interpretation von Brahms' Ungarischem Tanz Nr. 5. Dieses Stück hat die musikalische Form  $A_1A_2B_1B_2CA_3B_3B_4D$ , bestehend aus drei sich wiederholenden  $A$ -Teilen  $A_1$ ,  $A_2$  und  $A_3$ , aus vier sich wiederholenden  $B$ -Teilen  $B_1$ ,  $B_2$ ,  $B_3$  und  $B_4$ , sowie einem Mittelteil  $C$  und einem kurzen Schlussteil  $D$ . Die erste Instanz des Plug-Ins in Abb. 4 zeigt drei Zeilen an. Die erste Zeile (mit zwei Blöcken) entspricht der Wiederho-

lung  $A_2B_1B_2$  und  $A_3B_3B_4$ . Die zweite Zeile (mit vier Blöcken) repräsentiert die vier  $B$ -Teile und die dritte Zeile (mit drei Blöcken) die drei  $A$ -Teile. Die zweite Instanz des Plug-Ins von Abb. 4 zeigt die Wiederholungsstruktur auf einer feineren Detailstufe. So besteht zum Beispiel jeder  $A$ -Teil aus zwei harmonisch ähnlichen, sich wiederholenden Unterabschnitten. Die entsprechende Wiederholungsstruktur wird durch die sechs Blöcke der dritten Zeile repräsentiert. Durch Selektion eines bestimmten Blocks wird die Wiedergabe der zugehörigen Stelle innerhalb der Audioaufnahme gestartet beziehungsweise fortgesetzt. Unter Verwendung des Plug-Ins kann hier beispielsweise beliebig zwischen den  $A$ -Teilen hin- und hergesprungen werden. Im Hinblick auf musikwissenschaftliche Untersuchungen gestattet es diese Art der Navigation (*Intra-Dokument-Navigation*) allgemeiner, die unterschiedlich interpretierten Wiederholungen eines bestimmten Themas auf bequeme Weise auf Unterschiede hin zu untersuchen.

## 6 Komponente IV: AudioMatching

Eine der vom SyncPlayer unterstützten Sucharten ist die inhaltsbasierte Suche anhand kurzer gegebener Audioabschnitte unter Verwendung des Audiomatchings [Kurth and Müller, 2008]. Hierbei kann der Benutzer in einer gegebenen Audioaufnahme, visualisiert durch die zugehörige Wellenformdarstellung (vgl. Abb. 5), einen Ausschnitt markieren und diesen als Anfrage an die am Server vorliegende Datenbank übergeben. Das Suchergebnis besteht aus denjenigen Audioaufnahmen, die einen zur Anfrage ähnlichen Harmonieverlauf beinhalten. In einer Trefferliste werden die Trefferstellen innerhalb der gefundenen Audioaufnahmen angezeigt und sind anwählbar, so dass der Benutzer durch Anklicken einer solchen Trefferstelle die Wiedergabe direkt an entsprechender Position starten kann.

Als Beispiel wurden in Abb. 5 als Anfrage die ersten 20 Sekunden einer Bernstein-Interpretation von Beethovens Fünfter Symphonie in der Wellenformdarstellung selektiert. Dies entspricht hier genau dem ersten Vorkommen des bekannten Themas dieser Symphonie. Die Datenbank enthält fünf verschiedene Interpretationen von Beethovens Fünfter Symphonie: vier Orchesterversionen, aufgeführt von Bernstein, Karajan, Kegel, und Sawallisch sowie eine Klavierversion, gespielt von Scherbakov. Das erste Thema erscheint in der Exposition, in einer Wiederholung und schließlich noch einmal — in abgewandelter Form — in der Reprise. Damit sind 15 semantisch korrekte Treffer zu erwarten. Unterhalb der Wellenformdarstellung (Listenansicht) sind die Treffer aufgelistet, die das System bezüglich der Anfrage gefunden hat. Es wurden hier alle in der Datenbank enthaltenen Interpretationen gefunden. Das untere Fenster zeigt die einzelnen Treffer in den verschiedenen Interpretationen durch rote Blöcke, die unterhalb der zu einer Interpretation gehörigen Zeitleiste angebracht sind, an. Man kann erkennen, dass das System 14 Treffer gefunden hat, die auch alle im obigen musikalischen Sinne korrekt sind. Lediglich die Wiederholung des Themas in der Reprise der Klavierversion wurde nicht gefunden.

## 7 Komponente V: LyricsSeeker und KaraokePlayer

Eine weitere Suchart im SyncPlayer ist die Volltextsuche in den zu den Musikstücken gehörigen Gesangstexten. Hierzu bietet das LyricsSeeker-Plug-In eine Benutzerober-

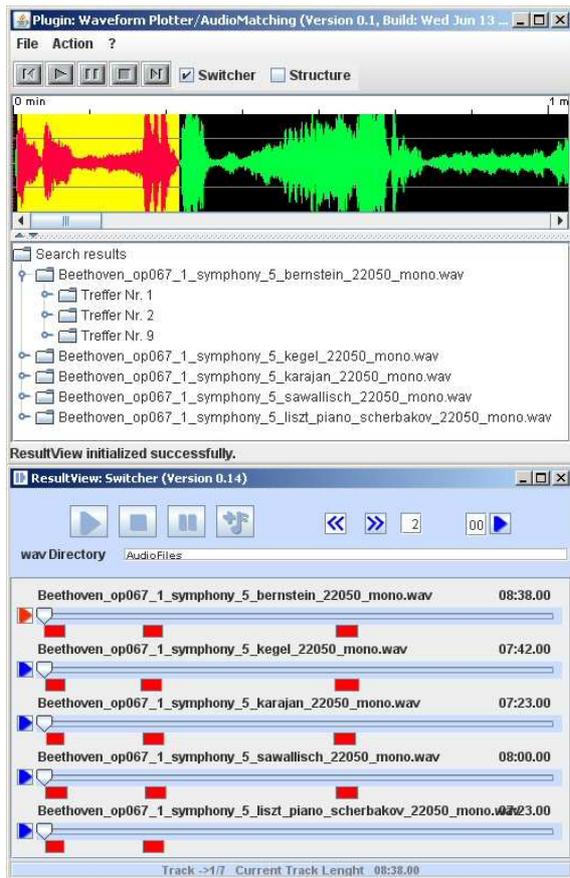


Abbildung 5: AudioMatching-Plug-In des SyncPlayers. **Oben:** Die Anfrage besteht aus einem Audioabschnitt (heller Hintergrund), welcher innerhalb einer Wellenformdarstellung ausgewählt werden kann. **Mitte:** Alle Audioaufnahmen, die Treffer zu dieser Anfrage enthalten, werden inklusive aller Trefferstellen aufgelistet. **Unten:** Benutzerschnittstelle zur Navigation in den die Treffer enthaltenden Audioaufnahmen. Die Funktionalität ist ähnlich zum AudioSwitcher, siehe Abb. 3. Zusätzlich werden die Treffer durch rote Blöcke repräsentiert, die durch Anklicken eine direkte Navigation zu den jeweiligen Trefferstellen erlauben.

fläche zur Formulierung von textuellen Anfragen sowie zur Präsentation von Suchergebnissen an. Neben der rein textuellen Präsentation der gefundenen Musikstücke, in denen der angefragte Text gesungen wird, bietet das Plug-In zusätzlich die Möglichkeit, zu jeder textuellen Trefferstelle den entsprechenden Audioausschnitt anzuhören.

In Abb. 6 ist auf der rechten Seite das LyricsSeeker-Plug-In dargestellt. In das obere Suchfeld können ein oder mehrere Suchbegriffe eingegeben und mittels des „Search“-Knopfs angefragt werden. Nach Anfragebearbeitung werden die Suchergebnisse unterhalb des Suchfelds dargestellt. Für jeden Treffer werden als Überschrift Interpret und Titel des entsprechenden Musikstücks sowie die Zeitposition der Trefferstelle aufgeführt. Weiterhin wird ein Textauszug aus dem Gesangstext angezeigt, in dem die Suchbegriffe vorkommen. Die darin enthaltenen Suchbegriffe werden zur besseren Orientierung hervorgehoben dargestellt. Durch einen Klick auf die zu einem Treffer gehörige



Abbildung 6: LyricsSeeker-Plug-In des SyncPlayers mit drei weiteren Instanzen einer Visualisierungskomponente zur karaokeartigen Anzeige von Gesangstext. Der Lyrics-Seeker gestattet eine Volltextsuche auf den Gesangstexten. Im Suchfeld (oben) wird die Anfrage formuliert. In der Ergebnisanzeige (unten) werden die Treffer untereinander dargestellt.

Überschrift wird die zu dem Musikstück gehörige Audioaufnahme in den SyncPlayer geladen und ab der Stelle, an der die Anfrage gesungen wird, wiedergegeben. Zusätzlich zur akustischen Wiedergabe wird der Gesangstext im Stil einer Karaoke-Anwendung synchronisiert angezeigt. Genaue wird dabei der umgebende Text zu dem in der Audioaufnahme aktuell gesungenen Wort angezeigt (vgl. Abb. 6, linke Seite); das aktuell gesungene Wort selbst wird durch einen laufenden Punkt unterhalb des Worts sowie eine veränderte Farbe hervorgehoben. Desweiteren lassen sich gleichzeitig mehrere Instanzen der Visualisierungskomponente zur karaokeartigen Anzeige von Gesangstexten anzeigen. Als praktisches Anwendungsbeispiel hierzu sei die Darstellung von Übersetzungen von Gesangstexten in andere Sprachen genannt. Auch die gleichzeitige Anzeige der Romanisierung schriftzeichenorientierter Sprachen kann hier, etwa im Rahmen von Lernumgebungen, verwendet werden.

## 8 Zusammenfassung und Ausblick

Das in dieser Arbeit im Überblick vorgestellte SyncPlayer-Framework bietet — aufbauend auf Mechanismen zur automatisierten Musikerschließung — Nutzerschnittstellen zur multimodalen Wiedergabe, Browsing und Suche in Musikkollektionen. Die in den einzelnen Abschnitten vorgestellten Komponenten können beliebig miteinander kombiniert werden, was eine neue Art des Musikerlebens gestattet. Diesbezüglich sei auf Abb. 7 verwiesen. Hier wird das Klavierlied „Frühlingstraum“ aus der „Winterreise“ von Schubert (D 911, Nr. 11), interpretiert von Thomas Allen, auf vielfältige, multimodale Weise wiedergegeben. Neben der akustischen Wiedergabe werden synchronisiert dazu der Gesangstext, die Klavierwalzendarstellung der Partitur (einmal Klavier mit Gesangsstimme, einmal nur Gesangsstimme) und die Wellenformdarstellung angezeigt. Ferner gestatten die beiden Strukturanzeigen auf verschiedenen Detailstufen eine bequeme Navigation zu bestimmten Teilen des Musikstücks.

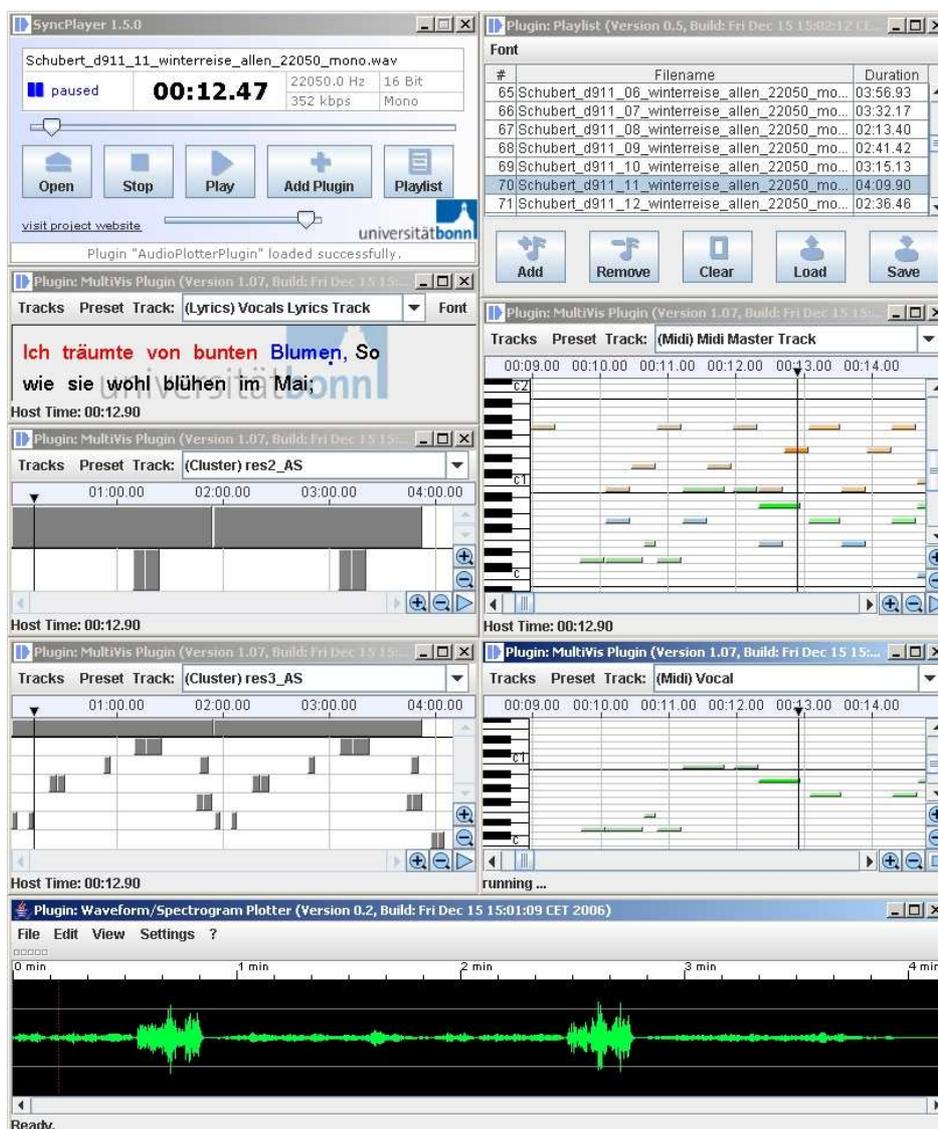


Abbildung 7: SyncPlayer mit verschiedenen Plug-Ins zum Klavierlied „Frühlingstraum“ aus der „Winterreise“ von Schubert (D 911, Nr. 11), interpretiert von Thomas Allen. Die Abbildung zeigt den SyncPlayer (links oben), die Playlist (rechts oben) und sechs Visualisierungs- und Navigations-Plug-Ins zur synchronisierten Anzeige des Gesangstexts, der Klavierwalzendarstellung und der Wellenformdarstellung sowie der Navigation in der musikalischen Struktur des Musikstücks.

Durch die Integration aktueller Forschungsergebnisse aus dem Bereich des MIR mit innovativen Nutzerschnittstellen leistet das SyncPlayer-Framework einen ersten Schritt in Richtung auf ein umfassendes automatisiertes Content Management System für heterogene Sammlungen von Musikdokumenten. Das momentan zum Einsatz in der Bayerischen Staatsbibliothek entwickelte PROBADO-Musik-Repositoryum erweitert wesentliche Teile des SyncPlayer-Frameworks im Hinblick auf einen Praxiseinsatz im bibliothekarischen Geschäftsgang inklusive der Nutzbarkeit durch Endanwender in der Bibliothek [Kurth *et al.*, 2008]. Die Grundlagenforschung im ARMADA-Projekt wird sich komplementär hierzu darauf konzentrieren, die in den meisten Fällen bislang nur für eingeschränkte Klassen von Musik einsatzfähigen Erschließungswerkzeuge weiterzuentwickeln. Zu den hier relevanten Aufgaben zählt beispielsweise die partielle Musiksynchronisation [Müller and Appelt, 2008], die sich mit der Verlinkung von nur teilweise übereinstimmenden Musikdokumenten beschäftigt.

## Literatur

- [Arifi *et al.*, 2004] Vlora Arifi, Michael Clausen, Frank Kurth, and Meinard Müller. Synchronization of music data in score-, MIDI- and PCM-format. *Computing in Musicology*, 13, 2004.
- [Bartsch and Wakefield, 2005] Mark A. Bartsch and Gregory H. Wakefield. Audio thumbnailing of popular music using chroma-based representations. *IEEE Trans. on Multimedia*, 7(1):96–104, February 2005.
- [Cano *et al.*, 2002] Pedro Cano, Eloi Battle, Ton Kalker, and Jaap Haitsma. A Review of Audio Fingerprinting. In *Proc. 5. IEEE Workshop on MMSP, St. Thomas, Virgin Islands, USA*, 2002.
- [Clausen and Kurth, 2004] Michael Clausen and Frank Kurth. A Unified Approach to Content-Based and Fault Tolerant Music Identification. *IEEE Transactions on Multimedia*, 6(5), October 2004.

- [Damm, 2007] David Damm. Textbasierte Musiksuche im Rahmen des SyncPlayer-Frameworks. Master Thesis, Dept. of Computer Science, University of Bonn, 2007.
- [Diet and Kurth, 2007] J. Diet and F. Kurth. The Probado Music Repository at the Bavarian State Library. In *Proc. ISMIR, Vienna, Austria*, pages 501–504, September 2007.
- [Dixon and Widmer, 2005] Simon Dixon and Gerhard Widmer. Match: A music alignment tool chest. In *Proc. ISMIR, London, GB*, 2005.
- [Fremerey, 2006] Christian Fremerey. SyncPlayer – a Framework for Content-Based Music Navigation. Master Thesis, Dept. of Computer Science, University of Bonn, 2006.
- [Good, 2001] Michael Good. MusicXML: An internet-friendly format for sheet music. In *Proc. XML Conference and Exposition*, 2001.
- [Goto, 2003] Masataka Goto. A chorus-section detecting method for musical audio signals. In *Proc. IEEE ICASSP, Hong Kong, China*, pages 437–440, 2003.
- [Hu et al., 2003] Ning Hu, Roger Dannenberg, and George Tzanetakis. Polyphonic audio matching and alignment for music retrieval. In *Proc. IEEE WASPAA, New Paltz, NY*, October 2003.
- [Huber, 1999] David Miles Huber. *The MIDI manual*. Focal Press, 1999.
- [Klapuri and Davy, 2006] Anssi Klapuri and Manuel Davy, editors. *Signal Processing Methods for Music Transcription*. Springer, New York, 2006.
- [Krottmaier et al., 2007] Harald Krottmaier, Frank Kurth, Thorsten Steenweg, Hans-Jürgen Appelrath, and Dieter Fellner. PROBADO - A Generic Repository Integration Framework. In *Proceedings of the 11th European Conference on Digital Libraries*, September 2007.
- [Kurth and Müller, 2008] F. Kurth and M. Müller. Efficient Index-based Audio Matching. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):382–395, February 2008.
- [Kurth et al., 2005] F. Kurth, M. Müller, D. Damm, C. Fremerey, A. Ribbrock, and M. Clausen. SyncPlayer—An Advanced System for Content-based Audio Access. In *Proc. ISMIR, London, GB*, 2005.
- [Kurth et al., 2007] F. Kurth, M. Müller, C. Fremerey, Y. Chang, and M. Clausen. Automated Synchronization of Scanned Sheet Music with Audio Recordings. In *Proc. ISMIR, Vienna, Austria*, pages 261–266, September 2007.
- [Kurth et al., 2008] F. Kurth, D. Damm, C. Fremerey, M. Müller, and M. Clausen. A framework for managing multimodal digitized music collections. In *To appear in Proceedings of the 12th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2008)*, 2008.
- [Maddage et al., 2004] Namunu C. Maddage, Changsheng Xu, Mohan S. Kankanhalli, and Xi Shao. Content-based music structure analysis with applications to music semantics understanding. In *Proc. ACM Multimedia*, pages 112–119, New York, NY, USA, 2004.
- [Müller and Kurth, 2007] M. Müller and F. Kurth. Towards Structural Analysis of Audio Recordings in the Presence of Musical Variations. *EURASIP Journal on Applied Signal Processing*, 2007(Article ID 89686):18 pages, January 2007.
- [Müller and Appelt, 2008] M. Müller and D. Appelt. Path-constrained partial music synchronization. In *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2008)*, to appear, 2008.
- [Müller et al., 2007] M. Müller, F. Kurth, D. Damm, C. Fremerey, and M. Clausen. Lyrics-based Audio Retrieval and Multimodal Navigation in Music Collections. In *Proceedings of the 11th European Conference on Digital Libraries*, September 2007.
- [Müller, 2007] Meinard Müller. *Information Retrieval for Music and Motion*. Springer, 2007.
- [Pardo, 2006] Bryan Pardo. Music information retrieval. *Special Issue, Commun. ACM*, 49(8):28–58, 2006.
- [Peeters et al., 2002] Geoffroy Peeters, Amaury La Burtche, and Xavier Rodet. Toward automatic music audio summary generation from signal analysis. In *Proc. ISMIR, Paris, France*, 2002.
- [Raphael, 2004] Christopher Raphael. A hybrid graphical model for aligning polyphonic audio with musical scores. In *Proc. ISMIR, Barcelona, Spain*, 2004.
- [Selfridge-Field, 1997] Eleanor Selfridge-Field, editor. *Beyond MIDI: the handbook of musical codes*. MIT Press, Cambridge, MA, USA, 1997.
- [Soulez et al., 2003] Ferréol Soulez, Xavier Rodet, and Diemo Schwarz. Improving polyphonic and poly-instrumental music to score alignment. In *Proc. ISMIR, Baltimore, USA*, 2003.
- [Turetsky and Ellis, 2003] Robert J. Turetsky and Daniel P.W. Ellis. Force-Aligning MIDI Syntheses for Polyphonic Music Transcription Generation. In *Proc. ISMIR, Baltimore, USA*, 2003.
- [Typke et al., 2005] Rainer Typke, Frans Wiering, and Remco C. Veltkamp. A survey of music information retrieval systems. In *ISMIR*, pages 153–160, 2005.
- [Wang et al., 2004] Ye Wang, Min-Yen Kan, Tin Lay Nwe, Arun Shenoy, and Jun Yin. LyricAlly: automatic synchronization of acoustic musical signals and textual lyrics. In *MULTIMEDIA '04: Proc. 12th annual ACM international conference on Multimedia*, pages 212–219, New York, NY, USA, 2004. ACM Press.