Automatisierte Annotation von Audiodaten mittels Synchronisationstechniken

Meinard Müller, Frank Kurth, Michael Clausen
{meinard, frank, clausen}@cs.uni-bonn.de

Abstract: Oft liegt ein Musikstück sowohl als Partitur als auch in Form unterschiedlicher Audioaufnahmen vor. In diesem Beitrag beschreiben wir ein Synchronisationsverfahren, das die Audiodaten automatisch mit der vorliegenden Partiturinformation verlinkt. Weiterhin stellen wir das SyncPlayer-System vor, das die so generierten Annotationen zur multimodalen Darstellung von Audiodaten verwendet.

1 Einleitung

Moderne digitale Musikbibliotheken enthalten multimediale Dokumente in zahlreichen Ausprägungen und Formaten, die ein Musikwerk auf verschiedenen Ebenen semantischer Ausdruckskraft beschreiben. Man denke hier beispielsweise an CD-Aufnahmen diverser Interpreten (Audiodaten), Noten (Partiturdaten), MIDI-Daten oder Gesangstexte. Bei der Erstellung einer multimedialen Musikbibliothek, die unter anderem eine inhaltsbasierte Suche und Datenanalyse unterstützen soll, spielen eine umfassende Annotation und Verlinkung des Datenbestandes eine entscheidende Rolle. Da dies aufgrund der enormen Datenmassen manuell nicht zu bewerkstelligen ist, sind Methoden zur automatischen Generierung semantisch hochwertiger Annotationen von zentralem Interesse. Wichtige Kernfragen sind hierbei die automatische Extraktion und Detektion inhaltsbasierter Informationen (z. B. Noten, Gesangstext, Klangfarben) aus Audiodaten als auch das Auffinden von prägnanter Strukturen (z. B. Akkordfolgen, Wiederholungen). In diesem Kontext spielt die sogenannte *Musiksynchronisation* eine wichtige Rolle, bei der es um die automatische Verlinkung von Daten unterschiedlicher Formate geht, siehe Abb. 1. Hierbei wird gerade das Vorliegen ein und desselben Musikstücks in *mehreren* Ausprägungen und Formaten ausgenutzt, um die automatisierte Inhaltserschließung von Audiodaten zu unterstützen.

2 Audio-Partitur-Synchronisation

In diesem Beitrag studieren wir das Szenario, in dem ein Musikstück sowohl als CD-Aufnahme (Audio) als auch in einem symbolischen Notenformat (Partitur) vorliegt. Unter einer *Audio-Partitur-Synchronisation* verstehen wir dann ein Verfahren, das zu einer bestimmten Position im Audiodatenstrom die entsprechende Stelle in der Partitur bestimmen kann. In diesem Sinne kann eine Audio-Partitur-Synchronisation als automatisierte Anno-

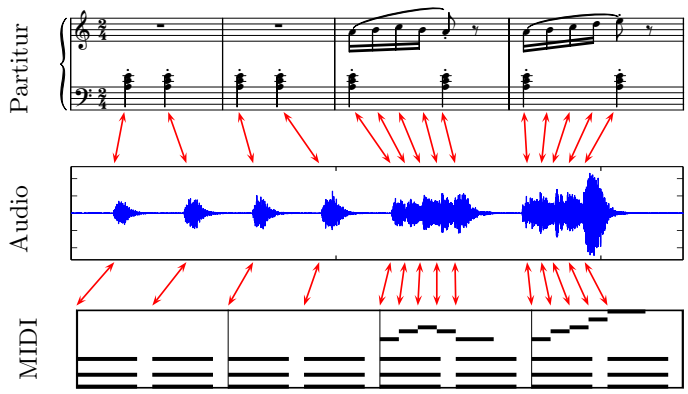

Abbildung 1: Verlinkung von Musikdaten in unterschiedlichen Formaten (Partitur, Audio, MIDI), die dasselbe Musikstück (die ersten vier Takte der Etüde Nr. 2, op. 100, F. Burgmüller) repräsentieren.

tation des Audiodatenstroms durch die Noten der Partitur oder auch als Extraktion bzw. Lokalisation von Noteninformation im Audiodatenstrom unter Ausnutzung des Vorwissens der Partiturdaten angesehen werden.

Da sich die rein symbolischen Partiturdaten grundlegend von den wellenformbasierten Audiodaten unterscheiden, stellt sich die Audio-Partitur-Synchronisation als ein schwieriges Problem dar. Auf der einen Seite besteht die Partitur aus notenbasierten Parametern wie Tonhöhen, Einsatzzeiten und Tonlängen, welche großen interpretatorischen Spielraum hinsichtlich des Tempos, der Dynamik oder der Ausführung von Notengruppen wie Trillern zulassen. Auf der anderen Seite kodiert eine CD-Aufnahme alle Parameter, die zur Rekonstruktion der akustischen Realisation (Wellenform) benötigt werden – die zugrundeliegenden Notenparameter sind allerdings nicht explizit gegeben. Daher gehen die meisten bisherigen Ansätze zur Audio-Partitur-Synchronisation in zwei Schritten vor: In einem ersten Schritt werden aus dem Audiodatenstrom geeignete Parameter extrahiert, die einen Vergleich mit den Partiturdaten erlauben. Im zweiten Schritt wird dann eine optimale Zuordnung mittels dynamischer Programmierung (DP) unter Verwendung geeigneter lokaler Ähnlichkeitsmaße berechnet. Für Details und weitere Literaturhinweise verweisen wir auf [1, 3, 4, 5]. Der Arbeit [3] folgend gehen wir nun auf einige Grundideen genauer ein.

DP-basierte Algorithmen, wie sie im Verlinkungsschritt eingesetzt werden, weisen ein quadratisches Laufzeitverhalten in der Eingabegröße auf und stellen daher meist den Flaschenhals bei der Audio-Partitur-Synchronisation dar. Daher verwenden wir in unserem Verfahren eine kleine Anzahl von semantisch ausdrucksstarken Merkmalen, die sowohl effizient aus dem Audiosignal extrahiert werden können als auch eine hohe Zeitauflösung aufweisen, wie sie im Hinblick auf eine präzise Synchronisation wichtig ist. Hierzu wird das Audiosignal unter Verwendung fortgeschrittener Filtertechniken (Filterbank aus elliptischen IIR-Filtern) gemäß den Klaviertönen in 88 Bänder zerlegt. Mittels energiebasierter Verfahren werden dann für jedes Band Kandidaten für Einsatzzeiten berechnet.

Im Fall polyphoner Musik stellt die Extraktion von Notenparametern ein extrem schwie-

riges Problem dar. Selbst für die Klasse polyphoner Klaviermusik, auf die wir uns im folgenden beschränken, bereiten z. B. Obertöne, Resonanz- und Schwebungseffekte, Vermischung von Klangspektren (verursacht durch das Haltepedal) oder auch das Vorliegen starker inharmonischer Komponenten (verursacht durch den Tastenanschlag) große Schwierigkeiten. Auch wenn die extrahierten Merkmale in Hinblick auf eine *Musiktranskription* unzureichend sein mögen, ermöglichen sie dennoch im allgemeinen eine ausgezeichnete *Musiksynchronisation*.

Nach einer geeigneten Aufarbeitung und Kodierung der Partiturdaten wird nun im zweiten Schritt mittels DP eine kostenoptimale zeitliche Verlinkung zwischen den Partitur- und Extraktionsparametern berechnet. Hierbei verwenden wir ein Verlinkungsmodell, welches sich von klassischen auf „dynamic time warping“ (DTW) basierenden Methoden, siehe z. B. [5], unterscheidet. Um eventuellen Unstimmigkeiten zwischen dem Partitur- und Audiodatenstrom, bedingt z. B. durch interpretatorische Abweichungen oder fehlerhafte Extraktion, Rechnung zu tragen, erzwingen wir nicht die Zuordnung aller Partitur- bzw. Extraktionsparameter, sondern erlauben auf beiden Seiten auch unverlinkte Ereignisse – ganz nach dem Motto: „Besser keine Zuordnung als eine schlechte Zuordnung.“ Darüber hinaus lassen wir uns bei der Definition des lokalen Ähnlichkeitsmaßes von folgendem einfachen aber weitreichenden Prinzip leiten: Die Partitur gibt uns vor, wonach im Audiodatenstrom zu suchen ist. Bei der Verlinkung werden also nur Extraktionsparameter berücksichtigt, die sich in der Partitur widerspiegeln. Für die technischen Details verweisen wir auf [3].

3 Experimente

Unser Verfahren wurde in MATLAB implementiert und anhand zahlreicher Beispiele polyphoner Klaviermusik unterschiedlicher Komplexität getestet, einschließlich Burgmüllers Etüden op. 100, Chopins Etüden op. 10 und einiger Klaviersonaten von Beethoven. Zur Demonstration und Bewertung der Synchronisationsergebnisse wurden diese *sonifiziert*. Hierzu sei daran erinnert, dass ein Audio-Partitur-Synchronisationsergebnis einer Zuordnung der musikalischen Einsatzzeiten der Partiturnoten mit den physikalischen Einsatzzeiten der entsprechenden Ereignisse im Audiodatenstrom entspricht. Für jede verlinkte Partiturnote wurde nun ein kurzer Sinuston der vorgegebenen Tonhöhe generiert, wobei die physikalische Einsatzzeit entsprechend der im Synchronisationsschritt ermittelten Zuordnung gewählt wurde. Schließlich wurde ein Stereo-Datenstrom erzeugt, welcher im linken Kanal eine Mono-Version des Audiodatenstroms und im rechten Kanal den Sinus-generierten Datenstrom enthält. Die so erzeugte Sonifikation[1] der Resultate zeigt, dass unser Verfahren für die eingeschränkte Musikklasse polyphoner Klaviermusik gute Synchronisationsergebnisse hoher Auflösung erzielt, die für Anwedungen wie die inhaltsbasierte Musiksuche oder zum Zwecke der zeitgleichen Notendarstellung beim Abspielen einer CD-Aufnahme mehr als ausreichend sind. Selbst plötzliche Tempoänderungen, ritardandi, accelerandi oder Fermaten konnten im allgemeinen gut erfasst werden.

[1] Sonifikationsergebnisse sind verfügbar unter `www-mmdb.iai.uni-bonn.de/download/sync/`.

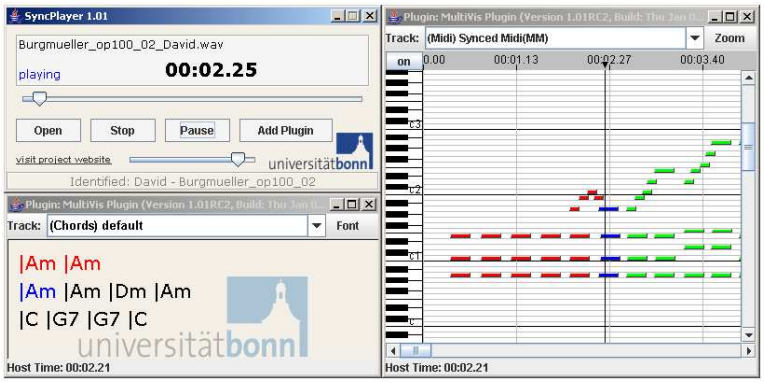

Abbildung 2: Benutzeransicht des SyncPlayer Systems: Bedieneinheit (links oben) mit Visualisierungsmodulen zur synchronen Darstellung von Partiturinformation (in Klavierwalzendarstellung, rechts) und Akkordfolgen (textuell, links unten).

4 Das SyncPlayer System

Die Ergebnisse des Synchronisationsschritts können in vielerlei Szenarien genutzt werden. Ein besonders interessantes Anwendungsgebiet ist dabei die synchrone Visualisierung der verlinkten Musikdaten während der akustischen Wiedergabe des zugehörigen Audiosignals. In diesem Abschnitt stellen wir das hierzu entwickelte SyncPlayer-System vor[2].

Die Grundfunktionalität des SyncPlayers umfasst die akustische Wiedergabe von Audiodateien (`WAV`- und `MP3`-Format), die auf einem lokalen Computersystem vorliegen. Wird eine solche Audiodatei im SyncPlayer geöffnet, versucht das System zunächst, die Audiodatei anhand eines digitalen Fingerabdrucks zu identifizieren. Hierzu verbindet sich das System mit einem zentralen Serverrechner, der über eine geeignete Datenbasis digitaler Fingerabdrücke verfügt. Zudem kann der Serverrechner auf die zuvor berechnete Verlinkungsinformationen zwischen den Fingerabdrücken und verfügbaren Musikdaten (z. B. Audio, Partitur, Gesangstext, Tabulatur) zurückgreifen. Glückt die Audioidentifikation, ermittelt das Serversystem die aktuelle Wiedergabeposition innerhalb des Audiostücks und liefert an den SyncPlayer (Client) zeitsynchron die verlinkten Musikdaten.

Der SyncPlayer verfügt zur Darstellung der Musikdaten über ein Visualisierungsmodul, das, je nach Art der zu einem Musikstück verfügbaren verlinkten Daten, parallel in verschiedenen Darstellungsmodi arbeiten kann. Abb. 2 zeigt die Benutzeransicht des SyncPlayers bei der Wiedergabe eines Stücks von Burgmüller, siehe Abb. 1. Zu diesem Stück liegen sowohl verlinkte Partiturdaten als auch verlinkte Akkorddaten vor. Der SyncPlayer zeigt in zwei Instanzen des Visualisierungsmoduls zeitsynchron zur akustischen Wiedergabe die entsprechenden Stellen in der textuellen Akkorddarstellung und in der Partitur in Form einer Klavierwalzendarstellung an. Eine ausführlichere Beschreibung des SyncPlayer-Systems findet sich in [2].

[2] Zum Download verfügbar unter `www-mmdb.iai.uni-bonn.de/projects/syncplayer/`.

Herkömmliche Systeme zur Musikwiedergabe wie etwa Audio- oder MIDI-Player verwenden jeweils nur *einen* Datentyp zur Erzeugung *aller* Parameter, die zur akustischen oder visuellen Wiedergabe benötigt werden. So wird z. B. die Audioausgabe in Programmen zur Bearbeitung der partiturnahen MIDI-Daten künstlich durch Synthesizer erzeugt. Im Gegensatz hierzu bietet unser SyncPlayer — in Verbindung mit Methoden zur Musiksynchronisation — die Möglichkeit zu einer angemesseneren multimodalen Musikdarstellung: Für jeden Wahrnehmungskanal wird die geeignetste Darstellungsform (akustische Darstellung durch Wiedergabe einer realen Audioaufnahme, visuelle Darstellung der Partiturdaten in Form einer Klavierwalze, Textdarstellung von Gesangspassagen) ausgewählt.

5 Ausblick

Die automatisierte Musikdatenerschließung stellt ein aktuelles Forschungsgebiet mit noch vielen ungelösten und interessanten Problemstellungen dar. Die Schwierigkeit liegt insbesondere in der Komplexität und Mannigfaltigkeit von Musikdaten begründet – nicht nur hinsichtlich unterschiedlichster Datenformate, sondern auch hinsichtlich der Gattung (z. B. Pop, Klassik, Jazz), der Instrumentation (z. B. Orchester, Klavier, Schlagzeug, Stimme) und vielen weiteren Parametern (z. B. Dynamik, Tempo, Klangfarbe). Für die Zukunft planen wir unter anderem, das Problem der Musiksynchronisation für allgemeinere Musikklassen effektiv und effizient zu lösen. Hierbei soll ein System entstehen, welches verschiedenartige, konkurrierende Strategien vereinigt, anstatt sich auf eine Strategie festzulegen. Der Synchronisationsalgorithmus kann weiterhin beträchtlich beschleunigt werden, falls im Vorfeld der eigentlichen DP-Berechnung schon eine kleine Anzahl *sicherer* Zuordnungen musikalischer und physikalischer Einsatzzeiten bekannt ist. Solche *Ankerkonfigurationen* können dann zur Umwandlung des globalen Synchronisationsproblems in eine Anzahl kleinerer, effizienter lösbarer Teilprobleme ausgenutzt werden, siehe [3].

Literatur

[1] Vlora Arifi, Michael Clausen, Frank Kurth, and Meinard Müller. Synchronization of Music Data in Score-, MIDI- and PCM-Format. In Walter B. Hewlett and Eleanor Selfridge-Fields, editors, *Computing in Musicology*, Volume 13, MIT Press, 2004.

[2] Frank Kurth, Meinard Müller, Andreas Ribbrock, Tido Röder, David Damm and Christian Fremerey. A Prototypical Service for Real-Time Access to Local Context-Based Music Information., Proc. of the *5th ISMIR*, Barcelona, Spain, 2004.

[3] Meinard Müller, Frank Kurth, and Tido Röder, Towards an Efficient Algorithm for Automatic Score-to-Audio Synchronization. Proc. of the *5th ISMIR*, Barcelona, Spain, 2004.

[4] Christopher Raphael, A Hybrid Graphical Model for Aligning Polyphonic Audio with Musical Scores Proc. of the *5th ISMIR*, Barcelona, Spain, 2004.

[5] Ferréol Soulez, Xavier Rodet and Diemo Schwarz, Improving polyphonic and poly-instrumental music to score alignment, Proc. of the *4th ISMIR*, Baltimore, Maryland, 2003.