# Audio Matching für symbolische Musikdaten

Frank Kurth, Meinard Müller, Christian Fremerey
*Institut für Informatik, Uni Bonn, Römerstr. 164, D-53117 Bonn, Email: {frank,meinard,fremerey}@iai.uni-bonn.de*

## Einleitung

In diesem Beitrag wird die Aufgabenstellung des Audio Matching auf den Fall symbolisch vorliegender Musikdaten (z.B. Partitur-, MIDI- oder durch Optical Music Recognition (OMR) gewonnene Daten) erweitert. Ausgangspunkt des klassischen Audio Matching Problems ist eine große Musikdatenbank, die typischer Weise mehrere verschiedene CD-Aufnahmen desselben Musikstücks enthält, wobei jede solche Aufnahme von unterschiedlichen Interpreten und in eventuell verschiedenen Besetzungen eingespielt wurde. Ist nun die Anfrage in Form eines kurzen Audioausschnitts einer bestimmten Interpretation gegeben, so sollen automatisch alle entsprechenden Ausschnitte in den anderen Interpretationen gefunden werden. Um nun auch symbolische Musikdaten sowohl untereinander als auch mit den Audiodaten vergleichbar zu machen, werden sowohl die Symbol- als auch die Audiodaten in eine gemeinsame "Mid-Level"-Darstellung transformiert. Hierzu verwenden wir Chroma-basierte CENS-Merkmale, die den groben Harmonieverlauf eines Musikstücks beschreiben. Weiterhin stellen wir ein auf den CENS-Merkmalen basierendes Matching-Verfahren vor, das einen hohen Grad an Robustheit gegenüber klanglichen und zeitlichen Variationen in den Musikdaten aufweist. Unser Matching-Verfahren bildet damit die Grundlage für robuste und praktikable Navigationstechniken in inhomogenen Musikdatenbeständen.

## Audio Matching

Wir fassen zunächst einen Lösungsansatz für einen Spezialfall des Audio Matching Problems zusammen, in dem sowohl die Musikdatenbank als auch die Anfrage aus CD-Aufnahmen bestehen [1]. In diesem Ansatz werden sowohl die Stücke der Datenbank als auch die Anfrage in Merkmalsfolgen über einem geeigneten Merkmalsraum transformiert. Das Audio Matching wird dann durch den Vergleich der Merkmalsfolgen durchgeführt.

Zur Merkmalsextraktion wird ein Audiosignal mittels einer Filterbank in 88 Tonhöhenbänder gemäß der temperierten Stimmung zerlegt. Für jedes Band wird durch geeignete Fensterung und Abtastratenänderung ein lokales Energiesignal mit einer Abtastrate von 10 Hz erzeugt. Anschließend werden alle zu gleichen Tonhöhenklassen gehörigen Bänder zu einem Chroma-Energiewert aufsummiert (z.B. korrespondieren hierbei die Bänder zu den Tonhöhen A0, A1,..., A7 zum Chroma-Band A). Hieraus ergeben sich Folgen von 12-dimensionalen Chroma-Vektoren. Zur Erhöhung der Robustheit gegenüber zeitlichen Verzerrungen werden diese Merkmale zusätzlich vektorweise quantisiert, zeitlich gefenstert, dezimiert und bzgl. ihrer Energie normalisiert. Die resultierenden 12-dimensionalen Merkmale werden als CENS-Merkmale (Chroma Energy Normalized Statistics) bezeichnet. Die Menge aller CENS-Merkmale bildet den hier verwendeten Merkmalsraum. Der linke Teil von Abb. 1 zeigt CENS-Merkmale für einen Ausschnitt des Klavierlieds D 911, Nr. 11 aus Schuberts *Winterreise* in einer Interpretation von Allen.

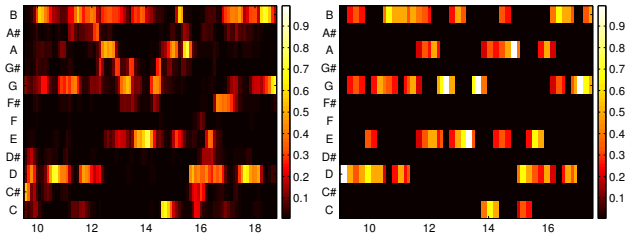


**Abbildung 1:** CENS-Merkmale für einen Ausschnitt (Takte 5–8) des Klavierlieds D 911, Nr. 11 aus Schuberts *Winterreise*. Links: CENS-Merkmale einer Interpretation von Allen, rechts: aus einer MIDI-Version gewonnene CENS-Merkmale.

Jedes Stück der Datenbank wird so in eine CENS-Folge mit einer Abtastrate von 1 Hz transformiert. Durch Konkatenation aller resultierender Vektorfolgen kann die Datenbank in eine einzige Folge $\mathcal{D} := (v^1, v^2, \ldots, v^N)$ von CENS-Merkmalen überführt werden. Zur Anfragebearbeitung wird ein kurzer (typischer Weise 10–30 Sekunden langer) Musikausschnitt ebenfalls in eine CENS-Folge $\mathcal{Q} := (w^1, w^2, \ldots, w^M)$ umgewandelt. Zum Audio Matching wird dann die Anfragefolge $\mathcal{Q}$ mit jeder Teilfolge von $\mathcal{D}$ verglichen, indem eine lokale Abstandsfunktion, $\Delta : [1 : N - M + 1] \to [0, 1]$, $\Delta(i) := 1 - \frac{1}{M} \sum_{m=1}^{M} \langle v^{i+m-1}, w^m \rangle$, ausgewertet wird. Die Stellen an denen $\Delta$ minimal wird, bilden in aufsteigender Reihenfolge die Menge der Trefferkandidaten. Durch Variation von Fensterbreite und Dezimierungsfaktor bei der Berechnung der CENS-Merkmale aus der Anfrage können unterschiedliche Wiedergabetempi simuliert werden, so dass beim Matching zusätzlich Stücke mit stark unterschiedlichen Geschwindigkeiten gefunden werden können, siehe [1].

## Symbolische Musikdatenformate

Um Audio Matching auf symbolischen Musikdaten durchführen zu können, diskutieren wir kurz die für uns relevanten Datenformate. In *Partiturformaten* können alle zur Darstellung einer Notenschrift benötigten Informationen erfasst werden. Beispiele hierfür sind proprietäre Formate von Musiknotationsprogrammen wie Capella und Finale, sowie das inzwischen weit verbreitete MusicXML-Format [2]. Charakteristisch für Partiturformate ist, dass hier lediglich uninterpretierte Notendaten

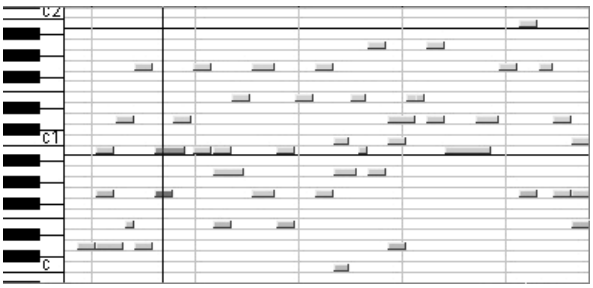

**Abbildung 2:** MIDI-Ausschnitt für das Schubert-Beispiel aus Abb. 1 in der der Klavierwalzendarstellung.

dargestellt werden.

Das MIDI-Format [3] stellt eine Art Zwischenformat dar. Hier werden sowohl symbolische Informationen über die Tonhöhen einzelner Noten und gewisse globale Partiturdaten (z.B. Tonart) eines Stücks, als auch interpretationsabhängige Informationen wie physikalische Einsatzzeiten und Tondauern kodiert. Abb. 2 zeigt den Ausschnitt einer MIDI-Datei zum obigen Schubert-Beispiel. Das MIDI-Fragment ist in der Klavierwalzendarstellung dargestellt, wobei jedes Rechteck eine Note repräsentiert, deren vertikale Position die diskrete Tonhöhe angibt. Linker Rand und horizontale Breite des Rechtecks repräsentieren jeweils Einsatzzeit und Tondauer der Note. Unter Annahme eines konstanten Wiedergabetempos können Partiturformate leicht nach MIDI konvertiert werden, so dass wir im folgenden davon ausgehen können, dass die symbolischen Musikdaten in diesem Format vorliegen. Aus Anwendungssicht ist dies besonders wichtig, da das Ausgabeformat von OMR-Software üblicherweise ein Partiturformat ist.

## Symbolisches Audio Matching

Die Übertragung der oben dargestellten Audio Matching Methode auf den Fall symbolischer Musikdaten erfolgt nun, indem die symbolischen MIDI-Daten in einem Vorverarbeitungsschritt in CENS-Merkmale überführt werden. Wir bemerken hierzu, dass die diskreten MIDI-Tonhöhen den bereits bei der Chroma-Berechnung verwendeten Tonhöhenbändern der temperierten Stimmung entsprechen. Aus einem MIDI-Dokument kann dann für jedes dieser Tonhöhenbänder ein diskretes Energiesignal konstruiert werden. Für jeden Zeitpunkt werden dabei genau diejenigen Energiesignale auf einen positiven konstanten Wert gesetzt, in deren Tonhöhenbändern laut MIDI-Dokument Noten aktiv sind. Alle anderen Energiesignale erhalten zu diesem Zeitpunkt den Wert 0. Anschaulich kann dieser Prozess so verstanden werden, dass die Klavierwalzendarstellung des zugehörigen MIDI-Dokuments (siehe Abb. 2) im wesentlichen die Indikatorfunktionen der konstruierten Energiesignale vorgibt.

Basierend auf den so gewonnenen lokalen Energiesignalen können nun analog zum Fall der CD-Aufnahmen Folgen von CENS-Merkmalen bestimmt werden. Im rechten Teil von Abb. 1 sind die so gewonnenen CENS-Merkmale für eine MIDI-Version obigen Ausschnitts aus Schuberts Winterreise dargestellt. Ein Vergleich mit den CENS-Merkmalen aus der Audioaufnahme zeigt deutlich die gemeinsame Grobstruktur. Insgesamt erhält man hierdurch für symbolische Audiodaten dieselbe Art von Merkmalen wie für CD-Aufnahmen und kann somit obige Matching Methoden verwenden.

## Experimente und Ausblick

Die vorgestellten Algorithmen zum Audio Matching für symbolische Musikdaten wurden in MATLAB implementiert. Hierzu wurden CENS-Merkmale für eine Datenbank aus 2.565 MIDI-Stücken mit klassischer Musik erstellt (entsprechend ca. $N = 736.000$ Merkmalen). Erste Studien zur Suche von MIDI-Anfragen in dieser MIDI-CENS-Datenbank zeigen, dass in der Regel Anfragen mit Längen von 10–15 Merkmalen ausreichen, um die angefragten Passagen in allen in der Datenbank enthaltenen MIDI-Interpretationen der jeweiligen Stücke aufzufinden. Weitere Experimente zeigen, dass auch eine Suche von MIDI-CENS-Anfragen in aus CD-Aufnahmen erzeugten CENS-Merkmalen sinnvoll ist. Hierzu wurde die in [1] vorgestellte CENS-Datenbank anhand ausgewählter Anfragen durchsucht. Ausführliche Evaluationsergebnisse werden an anderer Stelle veröffentlicht.

Zusammenfassend können unter Verwendung der in diesem Beitrag vorgestellten Methode signal- und symbolbasierte Musikformate vermöge eines gemeinsamen "Mid-Level"-Formats vergleichbar und wie aufgezeigt formatübergreifend durchsuchbar gemacht werden. Unsere Experimente deuten an, dass auch andere Aufgaben des Music Information Retrieval wie etwa die Musiksynchronisation durch die Ausnutzung eines solchen Mid-Level-Formats unterstützt werden können.

Die hier dargestellten Suchszenarien spielen im Kontext des aktuell von der Deutschen Forschungsgemeinschaft (DFG) geförderten Probado-Projekts [1] eine wichtige Rolle. Hier geht es unter anderem darum, in Bibliotheken vorliegende Sammlungen gescannter Partituren und digitalisierter Audioaufnahmen inhaltsbasiert durchsuchbar zu machen. Unsere Experimente zeigen, dass die vorgestellten CENS-Merkmale so robust sind, dass Suchaufgaben hier sogar auf mittels OMR aus den gescannten Partituren gewonnenen, teils fehlerbehafteten, Partiturdokumenten durchgeführt werden können.

## Literatur

[1] Meinard Müller, Frank Kurth, and Michael Clausen, Audio Matching via Chroma-Based Statistical Features. Proc. of the *6th ISMIR*, London, GB, 2005.

[2] MusicXML, URL: `http://www.recordare.com/xml.html`

[3] The MIDI Manufacturers Association, URL: `http://www.midi.org/`

[4] Probado – Bibliotheksdienste für allgemeine digitale Dokumente, URL: `http://www.probado.de`

[1] DFG-Förderung 554975 (1) Oldenburg BIB48 OLof 01-02