

Musicmatching bei Variabilitäten in der Harmonik und Polyphonie

Sebastian Ewert¹, Meinard Müller², Michael Clausen¹

¹ *Institut für Informatik III, Universität Bonn, Deutschland, Email: {ewerts,clausen}@iai.uni-bonn.de*

² *Universität des Saarlandes und MPI Informatik, Deutschland, Email: meinard@mpi-inf.mpg.de*

Einleitung

Im Jahr 1949 veröffentlichten Barlow und Morgenstern das *Dictionary of Musical Themes* [1], welches erstmals ermöglichte, ein Musikstück anhand eines dazu gehörigen Themas zu identifizieren. Voraussetzung dazu sind jedoch detaillierte musikalische Kenntnisse, weshalb erst eine Automatisierung dieses Prozesses auch musikalischen Laien die Benutzung erlaubt. In diesem Kontext sind *Musicmatching*-Methoden von besonderer Bedeutung. Ziel dieser Methoden ist, bei Anfrage eines kurzen Musikausschnitts (in diesem Fall monophone Themen im MIDI Format) alle hierzu musikalisch ähnlichen Ausschnitte innerhalb von Musikaufnahmen zu identifizieren. Entscheidend ist dabei der Begriff der musikalischen Ähnlichkeit. So liefern klassische Musicmatching-Methoden auch dann korrekte Ergebnisse, wenn sich die Anfrage und der zu identifizierende Ausschnitt in Klangfarbe, Instrumentierung oder Dynamik unterscheiden [3]. In dem hier vorgestellten Szenario ergeben sich aber darüber hinaus Unterschiede in Harmonik und Polyphonie, was mit klassischen Verfahren oftmals zu unbefriedigenden Resultaten führt.

In diesem Beitrag präsentieren wir erste Ergebnisse unserer Analysen, mit denen wir das Ziel verfolgen, die Robustheit klassischer Verfahren gegenüber Harmonie- oder Polyphonieunterschieden zu erhöhen. Im nächsten Abschnitt betrachten wir dazu zunächst, welche prinzipiellen, musikalisch begründeten Probleme sich aus der Aufgabenstellung ergeben, Aufnahmen anhand von Themen zu identifizieren. Im Anschluss werden mehrere Musicmatching-Methoden vorgestellt und deren Ergebnisse kurz diskutiert. Im letzten Abschnitt fassen wir die Resultate zusammen und geben einen Ausblick auf zukünftige Arbeiten.

Matching von Themen

Barlow und Morgenstern verzeichnen in [1] insbesondere Themen, die besonders prägnant für ein Stück sind. Dieser Umstand sollte bei der automatischen Identifizierung berücksichtigt werden. Wenn zum Beispiel ein Thema mehrfach in einem Stück gefunden werden kann, so gibt dies einen starken Hinweis darauf, dass das Stück korrekt durch das Thema identifiziert wird. Dabei ergibt sich jedoch das Problem, dass ein Thema typischerweise nicht in exakt gleicher Form wiederholt wird, sondern stets leicht modifiziert. In Abb. 1 wird dies anhand einiger Takte aus der Sonate Nr.1 von Beethoven illustriert. Das in [1] zu diesem Stück verzeichnete Thema findet sich in der rechten Hand der ersten beiden Takte, eine Variation des bekannten Mannheim-Themas. Dieses Thema wird mehrfach wiederholt, so in den Takten 3 und 4, 49 und 50 sowie 52 und 53. Obwohl die Wiederholungen in sehr ähnlicher Form aufzutreten scheinen, finden sich



Abbildung 1: Variationen des „Mannheim-Themas“ in den Takten 1-4 und 49-53 des ersten Satzes der Klaviersonate Nr.1 von Beethoven (Op. 2 Nr. 1 Satz 1).



Abbildung 2: Takte 1-4 des dritten Satzes der Klaviersonate Nr.12 von Beethoven (Op. 26 Satz 3).

dennoch deutliche Unterschiede. Zum Beispiel verändert sich zwischen Takt 1 und 3 nicht nur die Harmonie (f-Moll nach C-Dur), sondern auch die Melodie. Während im ersten Takt zwischen der ersten und zweiten Note drei Halbtonstufen liegen (F nach As), findet man an der gleichen Stelle im dritten Takt fünf Halbtonstufen (G nach C). Selbst aus musikwissenschaftlicher Sicht ist nicht im Allgemeinen definiert, wann eine Notenfolge eine modifizierte Variante eines gegebenen Themas darstellt. Alle Modifikationsmöglichkeiten beim Matching in robuster Weise zu berücksichtigen, stellt technisch somit eine große Herausforderung dar.

Neben der Problematik die Ähnlichkeit von Themen zu erfassen, stellt sich ferner auch die Frage, ob ein Thema ein Teil einer Menge von Noten ist. So lässt sich zum Beispiel argumentieren, dass in Abb. 2 sowohl das Thema Es-Es-As-Ces im ersten Takt vorkommt als auch das Thema Ces-As-Ces-As. Aus musikalischer Sicht sind beide Themen hier jedoch wenig repräsentativ für das Stück. Besonders prägnant treten Themen aber in der Melodie auf, weshalb man versuchen kann, die melodieunabhängigen, harmonietragenden Noten vorab von der Suche auszuschließen. Während bei Popmusik meist die lauteste oder die Gesangsstimme die Melodie bestimmt, gilt diese Annahme bei vielen klassischen Stücken jedoch nicht. Vielmehr existieren in einigen Werken sogar mehrere parallele Melodielinien, so dass aktuelle Methoden zur Melodieschätzung oftmals versagen.

Problematisch für ein robustes Matching ist ferner insbesondere die Länge der verzeichneten Themen, die oft nur aus fünf oder sechs Einzelnoten bestehen. Je kürzer ein

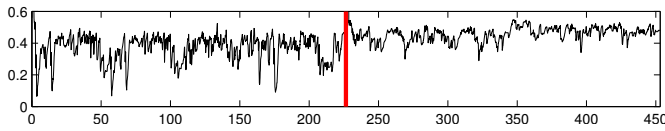


Abbildung 3: Konkatenierte Distanzkurve zu einer Aufnahme des ersten Satz der Beethoven Klaviersonate Nr.1 (Op. 2 Nr. 1 Satz 1). Angefragt wurde zuerst das „Mannheim-Thema“ (siehe Abb. 1), sowie ein weiteres Thema, das nicht in dem Stück auftritt. Ausschläge nach unten deuten auf lokale Ähnlichkeit hin.

Thema, desto wahrscheinlicher ist es, dass es in leicht modifizierter Form auch in Stücken zu finden ist, die nach [1] nicht mit dem Thema in Verbindung stehen.

Experimente

Aufgrund der soeben diskutierten prinzipiellen Probleme ist es interessant zu untersuchen, was aktuelle Matching-Methoden bei der Identifikation von Aufnahmen anhand von Themen leisten. Dabei wird das Matching von variierten Fassungen der Themen aufgrund der damit verbundenen Problematik hier nicht berücksichtigt. Zur Evaluierung verwenden wir eine Datenbank von Beethoven Klaviersonaten. Die Sonaten sind in insgesamt 101 Sätze aufgeteilt, für die jeweils eine Aufnahme einer Interpretation von Barenboim vorliegt. Daneben sind in [1] insgesamt 161 Themen zu den Sonaten verzeichnet, wobei zu jedem Satz mindestens ein Thema vorliegt.

Um nun ein automatisches Verfahren zur Identifikation der Werksbezeichnung der Aufnahmen zu entwickeln, betrachten wir im Folgenden mehrere Varianten eines klassischen Musicmatching-Verfahrens, für dessen Details wir auf [3] verweisen. Dabei ist zu beachten, dass im Vergleich zu [3] hier die Rollen von Anfrage und Datenbank vertauscht sind. Grundidee dieses Verfahrens ist, zunächst sowohl das Thema als auch die Aufnahme in eine Folge von CENS Merkmalsvektoren zu transformieren, welche die Intensität der Tonklassen $\{C, C\#, \dots, H\}$ über die Zeit beschreiben. Im nächsten Schritt kann dann die Merkmalsfolge des Themas lokal mit der der Aufnahme anhand einer Variante von Dynamic Time Warping (DTW) verglichen werden. Tempo- und Transpositionsunterschiede werden dabei automatisch ausgeglichen. Ergebnis ist eine Distanzkurve, deren Ausschläge nach unten auf eine starke Ähnlichkeit zwischen der Aufnahme und dem Thema hinweisen. Vergleicht man auf diese Weise eine zu identifizierende Aufnahme mit allen Themen, erhält man insgesamt 161 Distanzkurven, die im Anschluss zu einer Kurve konkateniert werden (siehe Abb. 3). Sucht man nun das Minimum in dieser Kurve, lässt sich feststellen, welches Thema der Aufnahme lokal am ähnlichsten ist. Über die mit dem Thema assoziierte Werksbezeichnung wird die Aufnahme schließlich identifiziert. Tabelle 1 zeigt in der ersten Zeile die Anzahl der Aufnahmen in Prozent, die auf diese Weise mit dem klassischen Verfahren korrekt identifiziert werden konnten. Es zeigt sich deutlich, dass das herkömmliche Verfahren in diesem Szenario ungeeignet ist.

In einem ersten Versuch werden deshalb die CENS Merkmale der Aufnahme modifiziert, indem jeweils die stärkste Komponente auf den Wert 1 und die übrigen auf

Tabelle 1: Evaluierungsergebnisse für verschiedene Varianten des klassischen Musicmatchings.

Merkmaltyp	%
CENS	15
CENS (dominanter Eintrag)	9
CENS (MIDI Melodie)	78
CENS (MultiF0)	48

den Wert 0 gesetzt werden. Die so modifizierten CENS Merkmale können als das Ergebnis eines sehr einfachen, intensitätsbasierten Melodieschätzers interpretiert werden. Die zweite Zeile von Tabelle 1 zeigt das Ergebnis der Kombination dieser modifizierten Merkmale und dem klassischen Verfahren. Wie sich erkennen lässt werden die Ergebnisse schlechter, was an der mangelnden Fähigkeit dieses einfachen Ansatzes liegt, die Hauptmelodie korrekt wiederzugeben.

In einem nächsten Experiment wurde untersucht, welche Ergebnisse mit einem guten Melodieschätzer zu erwarten sind. Da die Extraktion der Melodie aus Audio sehr schwer ist, werden in diesem Experiment MIDI Versionen der Beethoven Klaviersonaten identifiziert. Um aus diesen die Hauptmelodie zu schätzen, wurde eine modifizierte Version des Skyline Algorithmus verwendet. Dabei wird eine Note aus dem MIDI entfernt, wenn zu dieser Note parallel weitere Noten erklingen, deren Tonhöhe jedoch höher liegt. Die restlichen Noten korrelieren dann stark mit der Hauptmelodie (siehe [4]). Der Einsatz dieses Schätzers verbessert die Erkennungsrate deutlich, was darauf schließen lässt, dass die Themen sehr häufig in den Melodielinien auftreten. Diese Idee wurde in einem letzten Experiment auf Audio übertragen, wobei untersucht wurde, ob aktuelle Methoden zur Tonhöhenschätzung in Verbindung mit dem Skyline Algorithmus ähnliche Resultate liefern, wie der MIDI-basierte Melodieschätzer. Unter Einsatz eines Verfahrens von Klapuri [2] sinkt die Erkennungsrate zwar, dennoch bleiben die Ergebnisse deutlich über denen des klassischen Verfahrens.

Zusammenfassung und Ausblick

Die Identifikation von Audioaufnahmen anhand von monophonen Themen ist aufgrund diverser musikalisch bedingter Fragestellungen eine äußerst herausfordernde Aufgabenstellung. So konnte über klassische Musicmatching-Verfahren lediglich eine schwache Identifikationsleistung erzielt werden, die durch Integration von Melodieextraktionsverfahren jedoch deutlich erhöht werden konnte. Um die Identifikation weiter zu verbessern, werden wir in einem nächsten Schritt die Integration von Rhythmusextraktionsverfahren prüfen.

Literatur

- [1] H. Barlow and S. Morgenstern, *A Dictionary of Musical Themes*. Williams and Norgate, London, 1949.
- [2] A. Klapuri, “Multiple fundamental frequency estimation by summing harmonic amplitudes,” in *Proc. ISMIR*, 2006, pp. 221–216.
- [3] M. Müller, *Information Retrieval for Music and Motion*. Springer, 2007.
- [4] I. Suyoto, A. Uitdenbogerd, and F. Scholer, “Searching musical audio using symbolic queries,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 372–381, 2008.