

Notenschrift-Audio Synchronisation komplexer Orchesterwerke mittels Klavierauszug

Verena Thomas, Christian Fremerey, Sebastian Ewert, Michael Clausen

Institut für Informatik III, Universität Bonn, Deutschland

Email: {thomas, fremerey, ewerts, clausen}@iai.uni-bonn.de

Einleitung

Welcher Takt einer Partitur ist gerade in einer Einspielung des Musikstückes zu hören? Die Beantwortung dieser und ähnlicher Fragen ist Aufgabe der Notenschrift-Audio-Synchronisation. Einsatz finden solche Methoden beispielsweise zur Erleichterung der Navigation und Suche innerhalb von Musikstücken. Bisher wurden die entwickelten Verfahren für die Synchronisation von Partituren mit nicht zu komplexer Orchestrierung (hauptsächlich Klaviermusik) und entsprechenden Audioaufnahmen erprobt. Für komplexe Orchesterwerke ergeben sich bei der Notenschrift-Audio-Synchronisation jedoch neue Fragestellungen, von denen zwei in diesem Beitrag betrachtet werden sollen.

Zum einen stellen in diesem Kontext die in Orchesterwerken auftretenden transponierenden Instrumente einen kritischen Aspekt dar. Bei derartigen Instrumenten entsprechen die notierten Tonhöhen nicht den tatsächlich erklingenden. Liegen keine Informationen über die Transponierungen vor, so sind die betroffenen Stimmen relativ zur restlichen Partitur um mehrere Halbtöne verschoben. Im Zusammenhang mit der Notenschrift-Audio-Synchronisation ist zu betrachten, ob dieser Effekt Auswirkungen auf die zu erwartenden Ergebnisse hat.

Desweiteren existieren für viele Orchesterpartituren sogenannte Klavierauszüge, welche eine komprimierte Repräsentation des zugrundeliegenden Musikstückes darstellen. Es stellt sich die Frage, ob das Problem der transponierenden Instrumente umgangen werden kann, indem anstelle der Orchesterpartituren die Klavierauszüge mit voll instrumentierten Audioaufnahmen synchronisiert werden. Über die Taktaufteilung in den Noten kann anschließend die auf diese Weise berechnete Synchronisation auf die Orchesterpartitur übertragen werden.

In diesem Beitrag werden die Resultate einer Evaluation zu den oben aufgeführten Fragestellungen und Ansätzen vorgestellt. Dazu wurden für sieben Orchesterstücke auf verschiedene Arten die Synchronisationen einer Orchesterpartitur und eines Klavierauszuges mit einer Audioaufnahme berechnet. Im folgenden Abschnitt wird zunächst ein Überblick über das in den Experimenten verwendete Verfahren der Notenschrift-Audio-Synchronisation gegeben. Im Anschluss werden die berechneten Synchronisationen miteinander verglichen und bezüglich ihrer Qualität bewertet.

Notenschrift-Audio-Synchronisation

Musiksynchronisation im Allgemeinen hat das Ziel für jede Position in einer Repräsentation eines Musikstückes

(z.B.: Notenscan, MIDI, Audioaufnahme) die entsprechende Position in einer anderen Repräsentation zu bestimmen. Bei der Notenschrift-Audio-Synchronisation liegen als Ausgangsdokumente die Scans der Partitur sowie eine Audioaufnahme vor.

In einem ersten Schritt werden die beiden Musikdokumente in eine gemeinsame Darstellung überführt, durch die ein direkter Vergleich der musikalischen Inhalte möglich wird. Für Aufgaben wie Musiksynchronisation oder inhaltsbasierte Suche haben sich Chroma-basierte Merkmale bewährt. Zu Details über Chroma-Merkmale und deren Berechnung aus Audioaufnahmen siehe [5].

Für Notenscans wird zur Merkmalsextraktion zunächst über Optical Music Recognition (OMR) Verfahren der gegebenen Scan analysiert und dadurch die enthaltenen musikalischen Informationen (Notensysteme, Noten, Vorzeichen, etc.) erkannt. Bei dieser Analyse treten jedoch typischerweise Fehler auf (z.B.: nicht erkannte Vorzeichen, falsche Notendauern). Desweiteren ist bei Orchesterpartituren besonders kritisch, dass aktuelle OMR-Verfahren Transponierungen nicht erkennen. Die mittels OMR extrahierten musikalischen Inhalte werden anschließend in eine MIDI-Datei gewandelt. Aus dieser werden dann in ähnlicher Weise wie für Audioaufnahmen die Chroma-Merkmale berechnet.

Ausgehend von den berechneten Merkmals-Folgen für die beiden Musikdokumente, wird mittels eines Abstandsmaßes eine Ähnlichkeitsmatrix erstellt. Über Multiscale Dynamic Timewarping (MsDTW) [5] kann nun ein kostenminimaler Pfad durch diese Matrix berechnet werden, der die Synchronisation der beiden Musikdokumente kodiert.

Für Klavierwerke wurde von Ewert et al. [1] vorgeschlagen, den beschriebenen Prozess um eine weitere Merkmalsextraktion zu erweitern und dadurch die Synchronisationsresultate zu verbessern. Die gewählten Merkmale bilden die Noten-Einsatzzeiten in den Audioaufnahmen ab, wodurch eine genauere zeitliche Lokalisierung der Noten erreicht wird. Die für die beiden Merkmale berechneten Ähnlichkeitsmatrizen werden anschließend zu einer Matrix kombiniert. Über MsDTW wird danach der minimale Pfad durch diese Matrix berechnet. Grosche et al. [3] haben ein Verfahren zur Extraktion von Einsatzzeiten bei Orchestermusik vorgestellt (Novelty-Merkmale), durch das die Synchronisation für Orchesterwerke auf gleiche Weise verbessert werden konnte.

Experimente

Für die Experimente wurde ein Datensatz aus sieben Orchesterwerken zusammengestellt. Für jedes Mu-

sikstück lagen eine Orchesteraufnahme, eine Orchesterpartitur sowie ein Klavierauszug vor (Datenquellen: Bayerische Staatsbibliothek, RWC Music Database [2] und IMSLP/Petrucci Music Library [4]). Für drei der betrachteten Musikstücke sind über die RWC-Datenbank zusätzlich Beat-Annotationen der Audioaufnahmen verfügbar. Diese werden im Folgenden als Ground Truth für eine taktgenaue Synchronisation verwendet. Die OMR-Ergebnisse der Orchesterpartituren wurden manuell korrigiert und mit Transponierungsinformationen versehen. Für jedes Musikstück wurde die Audioaufnahme mit der um OMR-Fehler korrigierten und mit Transponierungen annotierten Orchesterpartitur (OP*), der unkorrigierten aber annotierten Orchesterpartitur (OP+), der unkorrigierten Orchesterpartitur (OP) und dem (unkorrigierten) Klavierauszug (KA) synchronisiert. Dabei wurden die manuell angegebenen Transponierungen während der Erstellung der MIDI-Dateien berücksichtigt und damit vor der Chroma-Berechnung ausgeglichen. Die Synchronisationen wurden sowohl mit dem MsDTW-Verfahren, als auch mit dem um die Novelty-Merkmale erweiterten Verfahren erstellt. Im Folgenden werden die Ergebnisse vorgestellt und diskutiert.

Ergebnisse mit Ground Truth

In Tabelle 1 werden die Ergebnisse für die Musikstücke mit vorliegender Ground Truth dargestellt. Die Werte geben die Abweichung der bei der Synchronisation bestimmten Taktanfänge in der Audioaufnahme zu den tatsächlichen Taktanfängen an. Wie zu erkennen ist, haben sowohl das Korrigieren der Orchesternoten als auch das Annotieren der Noten positiven Einfluss auf die Synchronisation. Eine weitere wichtige Einsicht ist, dass der Klavierauszug deutlich bessere oder ähnlich gute Ergebnisse liefert wie OP+ oder OP und nur minimal von den Ergebnissen für eine korrigierte und annotierte Orchesterpartitur abweicht.

		MsDTW		Novelty	
		mean	std	mean	std
OP*	Beethoven1	151	169	132	172
	Mozart1	76	74	51	52
	Mozart2	94	91	79	93
	Durchschnitt	80	84	67	79
OP+	Beethoven1	267	391	200	361
	Mozart1	76	74	51	52
	Mozart2	98	94	79	93
	Durchschnitt	110	140	83	127
OP	Beethoven1	462	700	494	957
	Mozart1	87	87	55	58
	Mozart2	98	96	83	102
	Durchschnitt	162	221	158	279
KA	Beethoven1	178	201	141	185
	Mozart1	89	86	52	55
	Mozart2	98	103	95	125
	Durchschnitt	91	98	72	91

Tabelle 1: Ergebnisse des taktweisen Vergleichs der Synchronisationen mit der aus den Beat-Annotation der RWC Music Library generierte Ground Truth (in ms).

Ergebnisse ohne Ground Truth

Für einige der Audioaufnahmen lagen keine Taktinformationen vor. Bei den entsprechenden Musikstücken wurde die für OP* berechnete Synchronisation als Vergleichsgrundlage verwendet. Damit verglichen wurden die Synchronisationsergebnisse für die unkorrigierte Orchesterpartitur und den Klavierauszug (siehe Tabelle 2). Hier ist ebenfalls klar erkennbar, dass das Fehlen der Transponierungsinformationen das zu erwartende Synchronisationsergebnis stark verschlechtern kann (siehe Beethoven5). Dagegen weicht der Klavierauszug deutlich geringer von der Synchronisation der korrigierten und annotierten Orchesterpartitur ab.

		MsDTW		Novelty	
		mean	std	mean	std
OP	Beethoven1	398	726	462	966
	Beethoven2	1097	1948	1004	1967
	Beethoven3	97	214	80	215
	Beethoven4	440	958	421	936
	Beethoven5	12047	5824	12160	5705
	Durchschnitt	2816	1934	2825	1958
KA	Beethoven1	86	139	48	119
	Beethoven2	183	529	83	173
	Beethoven3	156	359	161	435
	Beethoven4	90	157	75	169
	Beethoven5	312	557	402	675
	Durchschnitt	165	348	154	314

Tabelle 2: Ergebnisse des taktweisen Vergleichs der Synchronisation von OP* mit der Synchronisation von OP und KA. Es werden die Abweichungen von OP* in ms angegeben.

Resultat

Die vorgestellten Experimente zeigen, dass eine Synchronisation voll instrumentierter Audioaufnahmen mit einem passenden Klavierauszug möglich ist. Insbesondere zeigt sich, dass dies zu teilweise deutlich besseren Ergebnissen führt als die Synchronisation mit einer Orchesterpartitur ohne Kenntnis der transponierenden Instrumente. Für die Praxis bedeutet dies, dass im Vorfeld der Synchronisation von Orchesterwerken Informationen über transponierende Instrumente vorliegen sollten oder alternativ ein Klavierauszug zur Synchronisation verwendet werden kann. Über die Taktaufteilung in den Noten kann die für den Klavierauszug berechnete Synchronisation dann auf die Orchesterpartitur übertragen werden.

Literatur

- [1] S. Ewert, M. Müller, and P. Grosche, "High Resolution Audio Synchronization using Chroma Onset Features," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)*, 2009.
- [2] M. Goto, "RWC Music Database," Website, 2005, <http://staff.aist.go.jp/m.goto/RWC-MDB/>.
- [3] P. Grosche, M. Müller, and S. Ewert, "Combination of Onset-Features with Applications to High-Resolution Music Synchronization," in *Proceedings of the 35th International Conference on Acoustics (NAG/DAGA)*, 2009, pp. 357–360.
- [4] P. P. LLC, "IMSLP/Petrucci Music Library," Website, 2006, http://imslp.org/wiki/Main_Page.
- [5] M. Müller, *Information Retrieval for Music and motion*. Springer, 2007.