

Tempobasierte Segmentierung von Audioaufnahmen

Peter Grosche¹, Meinard Müller¹, Frank Kurth²¹ Saarland University and MPI Informatik, Campus E1.4, 66123 Saarbrücken, {pgrosche,meinard}@mpi-inf.mpg.de² Fraunhofer-FKIE, Abteilung KOM, 53343 Wachtberg-Werthhoven, kurth@fgan.de

Einleitung

Die automatische Segmentierung und Strukturierung ist für die automatisierte Verarbeitung von Musikaufnahmen von grundlegender Bedeutung. Das Ziel der Musiksegmentierung ist die Zerlegung eines Musikstücks in semantisch sinnvolle Abschnitte und elementare Einheiten. Diese Zerlegung eines Audiodatenstroms kann dann die Grundlage bilden für eine anschließende Weiterverarbeitung der Musikdaten, wie z.B. eine Klassifikation, Annotation oder Indexierung. Musikalisch sinnvolle Segmente eines Musikstücks zeichnen sich oft dadurch aus, dass sie in sich homogen hinsichtlich bestimmter musikalischer Aspekte sind. Darüber hinaus unterscheiden sich unterschiedliche Abschnitte häufig bezüglich ihrer Instrumentierung, der vorherrschenden Klangfarbe und rhythmischer, melodischer oder harmonischer Eigenschaften. In diesen Fällen bietet sich die Verwendung homogenitätsbasierter Segmentierungsverfahren an [1], die eine Unterteilung in Abschnitte vornehmen, die Homogenität bezüglich bestimmter Signaleigenschaften aufweisen. Hierzu werden Darstellungsformen der Musiksignale benötigt, auf der einen Seite semantische sinnvolle, musikalisch aussagekräftige Eigenschaften der Signale beschreiben und auf der anderen Seite robust gegenüber Abweichungen sind. Des Weiteren ist eine niedrige Dimensionalität gefordert um eine effiziente Verarbeitung zu ermöglichen. Zur Beschreibung der Klangfarbe erfüllen z.B. MFCC Merkmale diese Voraussetzungen ebenso, wie CHROMA Merkmale zur Beschreibung der harmonischen Eigenschaften eines Stücks [2].

In diesem Beitrag stellen wir Merkmale vor die das lokal vorherrschende Tempo eines Musikstücks beschreiben und die Anforderungen hinsichtlich Robustheit und Dimensionalität erfüllen. Des Weiteren diskutieren wir wie mit diesen Merkmalen eine tempobasierte Segmentierung vorgenommen werden kann indem das Musikstück in Abschnitte mit homogenem Tempo eingeteilt wird und zeigen wie dies eine sinnvolle Ergänzung zur Segmentierung hinsichtlich der Klangfarbe oder Harmonie eines Musikstücks darstellt.

Zyklisches Tempogramm

Verfahren zur automatischen Bestimmung des Tempos einer Musikaufnahme gehen im Allgemeinen in zwei Schritten vor. Im ersten Schritt werden aus dem Audiosignal Einsatzzeitpunkte von Noten extrahiert. Hierzu wird eine sogenannte *Novelty-Kurve* berechnet. Diese erfasst Änderungen bestimmter Signaleigenschaften und erlaubt es wahrscheinliche Zeitpunkte für Noteneinsätze zu selektieren [3]. Die zugrunde liegende Annahme ist,

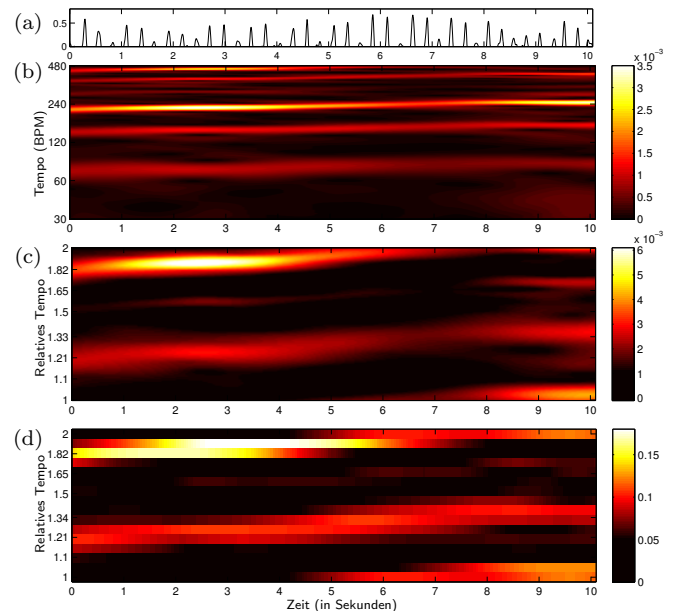


Abbildung 1: (a) Novelty-Kurve eines Ausschnitts eines Walzers aus der zweiten Jazz Suite von Schostakovich. (b) Tempogramm Darstellung abgeleitet aus (a). (c) Zyklisches Tempogramm abgeleitet aus (b). (d) Diskretisiertes zyklisches Tempogramm mit 15 Dimensionen.

dass der Beginn einer neuen Note mit einer Veränderung der Energie, der Tonhöhe oder der Klangfarbe des Signals einhergeht. Abb. 1(a) zeigt eine Novelty-Kurve für eine Audioaufnahme eines Walzers aus der zweiten Jazz Suite von Schostakovich, die Änderungen der spektralen Energieverteilung des Signals erfasst [4]. Die Maxima dieser Kurve deuten auf Zeitpunkte mit wahrscheinlichen Einsätzen von Noten hin.

Im zweiten Schritt wird nun das Tempo der Musikaufnahme bestimmt. Hierzu werden die Noteneinsatzzeitkandidaten hinsichtlich zugrunde liegender Periodizitäten analysiert. Zu diesem Zweck können bekannte Methoden der Signalverarbeitung wie die Fouriertransformation oder die Autokorrelation eingesetzt werden. Abb. 1(b) zeigt das Ergebnis einer gefensterten Fouriertransformation der Novelty-Kurve des Schostakovich Beispiels. Diese, auch *Tempogramm* genannte Darstellung beschreibt für jeden Zeitpunkt des Musikstücks das lokal vorherrschende Tempo in Schlägen pro Minute (BPM) und entspricht einem Spektrogramm der Novelty-Kurve, mit dem Unterschied, dass die physikalische Größe Frequenz (in Hz) hier als musikalische Größe Tempo (in BPM) interpretiert wird. So weist das Tempogramm in diesem Fall hohe Intensitäten im Bereich 210 bis 140 BPM auf. Dies entspricht dem musikalischen Tempo des Stücks in Viertelnoten pro Minute. Eine robuste Ableitung des musikalischen Tempos aus dieser Darstellung

ist allerdings nicht direkt möglich, da bedingt durch die harmonische Struktur und die verschiedenen metrischen Ebenen zusätzlich zu dem musikalischen Tempo weitere Einträge eine hohe Intensität aufweisen, wie z.B. bei 70, 140, 280, 350 und 420 BPM.

Um diese Unklarheiten und die daraus entstehenden Verwechslungen bei Temposchätzung zu umgehen wenden wir eine Strategie an, die vergleichbar ist mit der Berechnung von CHROMA Merkmalen [5]. Hierzu bezeichnen wir zwei Tempi τ_1 und τ_2 als *Oktav-äquivalent*, wenn $\tau_1 = 2^k \tau_2$ für ein beliebiges $k \in \mathbb{Z}$. Dies ergibt für ein Tempo $\tau = 120$ BPM die Tempo-Äquivalenzklasse $[\tau] = \{\dots, 30, 60, 120, 240, 480, \dots\}$. Unter diesen Äquivalenzen bilden wir nun das gesamte Tempogramm auf eine Referenzoktave ab indem die Intensitäten aller äquivalenten Tempi des Tempogramms aufsummiert werden. Diese Darstellung wird als *zyklisches Tempogramm* [4] bezeichnet, da die Tempo-Äquivalenzklassen in dem Sinne einer zyklischen Struktur unterliegen, als dass eine Änderungen des Tempos eines Musikstücks in einer zyklische Rotation des zyklischen Tempogramms resultiert.

Abb. 1(c) zeigt das zyklische Tempogramm für das Tempogramm aus Abb. 1(c) für die Referenzoktave 60 bis 120 BPM, d.h. das relative Tempo bezieht sich auf 60 BPM. Der hohe Eintrag im Bereich des relativen Tempo von 1.8 beschreibt folglich die Äquivalenzklasse $[\tau] = \{\dots, 55, 110, 220, 330, 440, \dots\}$. Unter diesen Oktav-Äquivalenzen erhält man somit eine niedrigdimensionale und sehr robuste Repräsentation des lokalen Tempos eines Musikstücks, die aber dennoch in der Lage ist sehr präzise und semantisch aussagekräftig die lokalen Charakteristiken eines Stücks zu beschreiben. Zur weiteren Reduzierung der Dimensionalität kann das zyklische Tempogramm quantisiert werden. Abb. 1(d) zeigt das zyklische Tempogramm aus Abb. 1(c) reduziert auf 15 dimensionale Merkmalsvektoren.

Tempobasierte Segmentierung

Im folgenden skizzieren wir, wie das zyklische Tempogramm für die automatische Segmentierung von Musikstücken Verwendung finden kann. Für die homogenitätsbasierte Segmentierung bieten sich automatische Cluster Algorithmen an, wie z.B. K-Means oder hierarchisch Verfahren, die die Merkmalsvektoren in Gruppen einteilen, die bestimmte Ähnlichkeiten aufweisen [1]. In vielen Fällen ist die eingangs erwähnte Homogenitätsannahme hinsichtlich der Klangfarbe oder der Harmonie einer Musikaufnahme nicht erfüllt. Hier kann die tempobasierte Segmentierung einen Beitrag zur Auffindung musikalisch sinnvoller Segmente leisten.

Abb. 2 zeigt verschiedenen Merkmalsfolgen für eine Klavierversion von Brahms Ungarischem Tanz Nr. 5. Dieses Stück ist geprägt von mehreren abrupten Tempowechseln und musikalischen Abschnitten die in unterschiedlichem Tempo gespielt werden. Insbesondere weist es einen Mittelteil auf, der erheblich langsamer gespielt wird als das übrige Stück. Abb. 2(c) zeigt eine exemplarische zwei Klassen Segmentierung des Musikstücks, die den Mittel-

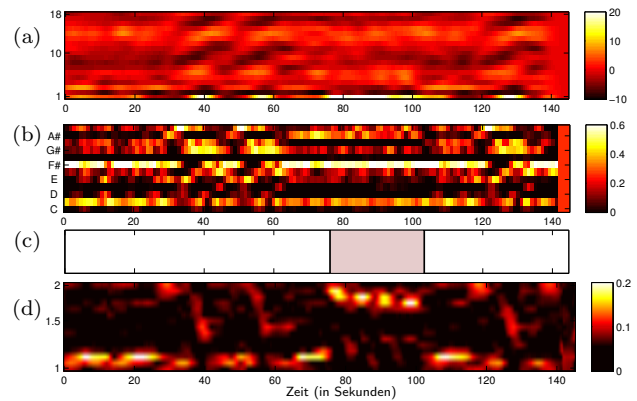


Abbildung 2: Vergleich verschiedener Merkmale für eine Klavierversion von Brahms Ungarischer Tanz Nr. 5 (a) MFCC Merkmale. (b) CHROMA Merkmale. (c) Tempobasierte Segmentierung. (d) Zyklisches Tempogramm.

teil des Stücks hervorhebt. Die MFCC (Abb. 2(a)) und auch CHROMA Merkmale (Abb. 2(b)) können dieses Segment nicht offenbaren da die Homogenitätsannahme bezüglich dieser Merkmale nicht erfüllt ist. So ändert sich die Instrumentierung an den Grenzen dieses Segments nicht, weshalb sich die Klangfarbe nur geringfügig ändert. Auch die harmonischen Eigenschaften dieses Segments unterscheiden sich nicht vom Rest des Stücks. Allerdings tritt im zyklischen Tempogramm (Abb. 2(d)) dieses Segment aufgrund des abweichenden Tempos sehr deutlich hervor und zeigt damit die Eignung dieses Merkmals für die Segmentierung von Musikaufnahmen.

Ausblick

Das zyklische Tempogramm kann musikalisch aussagekräftige Segmentierungen rein auf der Grundlage einer niedrigdimensionalen Beschreibung des lokal vorherrschenden Tempos unterstützen, die robust ist gegenüber Tempoverwechslungen. In der Praxis wird für eine robuste Segmentierung von Musikaufnahmen eine Kombination verschiedener Merkmale benötigt, um der Vielfältigkeit der Musik gerecht zu werden. Diese Eigenschaften des zyklischen Tempogramms erlauben darüber hinaus den Einsatz des Merkmals für weitere Aufgabe der automatischen Verarbeitung von Musikaufnahmen, wie z.B. einer inhaltsbasierten Suche, Strukturierung oder Indexierung von Datenbanken.

Literatur

- [1] M. Levy and M. B. Sandler. Structural segmentation of musical audio by constrained clustering. *IEEE Trans. on Audio, Speech and Language Processing*, 16(2):318–326, 2008.
- [2] K. Jensen. Multiple scale music segmentation using rhythm, timbre, and harmony. *EURASIP J. Applied Signal Processing*, 2007(1):159–159, 2007.
- [3] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. E. P. Davies, and M. B. Sandler. A tutorial on onset detection in music signals. *IEEE Trans. on Speech and Audio Processing*, 13(5):1035–1047, 2005.
- [4] P. Grosche, M. Müller, and F. Kurth. Cyclic tempogram – a mid-level tempo representation for music signals. In *Proc. of IEEE ICASSP*, Dallas, TX, USA, 2010.
- [5] Meinard Müller. *Information Retrieval for Music and Motion*. Springer, 2007.