

# Ein Baseline-Experiment zur Klassifizierung von Problemen bei der Akkorderkennung

Verena Konz<sup>1</sup>, Meinard Müller<sup>1</sup>, Sebastian Ewert<sup>2</sup>

<sup>1</sup> Universität des Saarlandes und MPI Informatik, Saarbrücken, Deutschland Email: {vkonz,meinard}@mpi-inf.mpg.de

<sup>2</sup> Institut für Informatik III, Universität Bonn, Deutschland Email: ewerts@iai.uni-bonn.de

## Einleitung

In den letzten Jahren hat die Akkorderkennung im Bereich der automatisierten Musikverarbeitung ein immer grösseres Forschungsinteresse erlangt [1, 2]. Dies ist nicht verwunderlich, stellt doch die Harmonie ein grundlegendes Prinzip westlicher Musik dar. So ist ein Musikstück wesentlich von seiner harmonischen Progression, der zeitlichen Abfolge verschiedener Akkorde, geprägt. Hierbei versteht man unter einem Akkord den Zusammenklang verschiedener Töne. Aufgrund dieser Basisfunktion der Harmonie für ein Musikstück, dienen Akkordfolgen ausserdem als Zwischen-Merkmal-darstellungen für andere Fragestellungen in der automatisierten Musikverarbeitung, wie z. B. die Segmentierung, die Indexierung oder die inhaltsbasierte Suche in Audiodatenbeständen [3].

In diesem Beitrag wird ein template-basiertes Baseline-Verfahren zur Akkorderkennung vorgestellt und auf dem Audiodatenbestand der zwölf Studioalben der Beatles ausgewertet. Hierbei steht eine Klassifizierung von Problemen im Vordergrund, die bei der Akkorderkennung auftreten können. In dem Zusammenhang wird explizit auf die Rolle des Tunings eingegangen und gezeigt, dass der Ausgleich von Tuningabweichungen einen wesentlichen Einfluss auf das Ergebnis der Akkorderkennung hat.

## Baseline-Verfahren zur Akkorderkennung

Im Folgenden wird das in den Experimenten verwendete Baseline-Verfahren zur Akkorderkennung skizziert. Zunächst extrahieren wir mittels einer Filterbank, die das Audiosignal in 88 Tonhöhenbänder zerlegt, Tonhöhenmerkmale. Diese werden in 12-dimensionale CENS-Merkmale (Chroma Energy Normalized Statistics) überführt, die als spezielle Tonhöhen-Merkmale verstanden werden können, und die Intensität der 12 Tonklassen  $\{C, C\sharp, D, D\sharp, \dots, H\}$  im Zeitverlauf angeben [4]. Wir erhalten somit 10 CENS-Vektoren pro Sekunde, wobei jeder CENS-Vektor grob gesagt 1100 ms des ursprünglichen Audiosignals miteinbezieht.

Nun definieren wir 24 verschiedene Akkord-Templates, 12 Dur-Templates und 12 Moll-Templates. Die betrachteten Templates sind 12-dimensionale Vektoren, in denen die 3 Komponenten des zugehörigen Dur-/Mollakkords auf 1 gesetzt werden und alle anderen Komponenten auf 0. Beispielsweise erhalten wir für C-Dur das Template  $(1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0)$  und für c-moll das Template  $(1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0)$ . Wir definieren nun eine Distanzfunktion  $d$ , die den Abstand des  $i$ -ten Merkmalsvektors  $y_i$  zu einem beliebig fest gewählten Template  $t$  bestimmt,  $d : [0, 1]^{12} \times [0, 1]^{12}, d(t, y_i) = 1 - \langle t, y_i \rangle$ , wobei

$\langle \cdot, \cdot \rangle$  das innere Produkt bezeichnet. Für die Akkordberechnung im  $i$ -ten Zeitfenster wird nun dasjenige Template ausgewählt, das den minimalen Abstand  $d$  zum  $i$ -ten Merkmalsvektor hat.

## Tuning

Für die Stimmung im Bereich westlicher Musik wird heutzutage meistens die gleichstufige Stimmung verwendet. Dennoch stellt man oft in Aufnahmen fest, dass die verwendete Stimmung von der Standardstimmung, bei der der Kammerton  $a$  mit 440 Hz identifiziert wird, geringfügig abweicht. Will man die Akkorde eines Musikstücks auf der Basis einer Audiodatei extrahieren, so ist es insofern sinnvoll den Aspekt des Tunings miteinzubeziehen. Zu diesem Zweck verschieben wir die zur Extraktion der Tonhöhen-Merkmale verwendete Filterbank auf die Art, dass wir 6 verschiedene Filterbänke erhalten, die Tuningabweichungen von 0, 0.25, 0.33, 0.5, 0.67 und 0.75 Halbtönen betrachten. In Kombination mit 12 zyklischen Shifts der Chromavektoren, die jeweils einer Semiton-Verschiebung entsprechen, erhalten wir somit 72 verschiedene Merkmalsversionen. Wir berechnen nun mittels des zuvor beschriebenen Baseline-Verfahrens F-Measure-Werte für die Akkorderkennung für alle 72 Merkmalsversionen und betrachten letztendlich diejenige Merkmalsversion, die den F-Measure Wert maximiert.

## Experimente

Ausgangspunkt für die Auswertung im vorliegenden Beitrag bildet der Datenbestand der zwölf Studioalben der Beatles, die insgesamt 180 Songs umfassen. Zu diesem Datenbestand liegen detaillierte Akkord-Annotationen vor, die im folgenden als Groundtruth dienen [2]. Da wir die Auswertung nur auf den 24 Dur- und Mollakkorden vornehmen, müssen die Akkordkürzel der Groundtruth zunächst auf den jeweils nächsten Dur bzw. Mollakkord abgebildet werden, bevor sie mit dem berechneten Resultat verglichen werden können. Die Auswertung im Folgenden erfolgt pro Zeitfenster.

Wir zeigen nun beispielhaft Probleme auf, die bei der Akkorderkennung auftreten können. Hierzu betrachten wir zunächst die ersten vier Takte des Songs *Let it be* (siehe Abb. 1 a)). In Abb.1 b) ist eine Visualisierung der Akkorderkennung für diesen Ausschnitt dargestellt. Hierbei sind auf der vertikalen Achse alle 24 Dur-/Mollakkorde aufgetragen. Hellgrüne Blöcke geben nun diejenigen Zeitpositionen und entsprechenden Akkorde an, die manuell annotiert wurden, also der Groundtruth entsprechen. An den Stellen, wo das berechnete Resultat mit der Groundtruth übereinstimmt, sind die grünen Bereiche

nächträglich blau markiert. Dort, wo das berechnete Resultat von der Groundtruth abweicht, sind sie rot. So kann man nun direkt aus der Visualisierung abgelesen, wo Probleme bei der Akkorderkennung vorliegen (rot/grün) und wo der berechnete Akkord mit der Groundtruth übereinstimmt (blau). So erkennt man, dass zu Beginn des Ausschnittes die Akkorderkennung gute Ergebnisse liefert. Die erste längere Fehlklassifikation geschieht im Bereich zwischen ca. 5 und 6 Sekunden. Hier wurde offensichtlich manuell F-Dur annotiert, allerdings in der berechneten Annotation stattdessen a-moll bestimmt. Diese Stelle entspricht in den Noten der 3. Viertel im 2. Takt. Hier liegt musikalisch ein F-Major-Seven-Akkord vor, d.h. ein F-Dur Akkord, bei dem eine grosse Septime hinzugefügt ist (hier das e). Der Akkord besteht somit aus den vier Tönen f, a, c und e. A-moll jedoch besteht aus den drei gemeinsamen Tönen a, c und e. Dies ist an dieser Stelle der musikalische Grund für die Fehlklassifikation. Ein ähnlicher Fall liegt in der zweiten Fehlklassifikation (bei ca. 6-7 Sekunden) vor, wo statt F-Dur d-moll erkannt wird. Dieses Problem, dass in der musikalischen Ambivalenz der Akkorde liegt, tritt oft in der Akkorderkennung auf. Selten liegen reine Dur- oder Mollakkorde vor; oftmals werden Töne hinzugefügt oder weggelassen, die eine eindeutige Zuordnung des Akkordes erschweren.

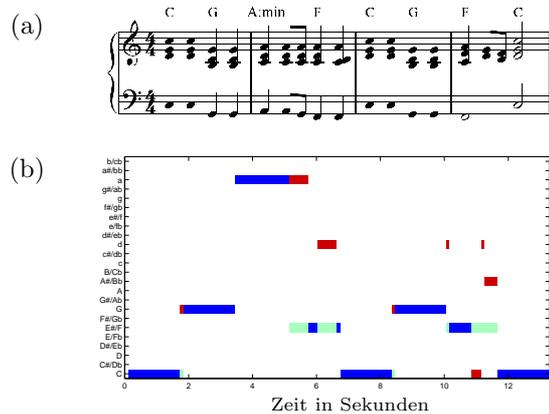
Ein weiteres zentrales Problem bei der Akkorderkennung ist das Tuning. Um diesen Aspekt zu untersuchen, haben wir zunächst das Baseline-Verfahren für alle 180 Songs ohne Berücksichtigung des Tunings angewendet. In einem zweiten Schritt haben wir den Tuningaspekt in das Baseline-Verfahren integriert. In Tab. 1 sehen wir eine Gegenüberstellung der F-Measure-Werte für eine Auswahl von Songs, bei denen das Tuning eine sehr erhebliche Rolle spielt. Hierbei bezeichnen  $F_{orig}$  bzw.  $F_{tune}$  den F-Measure-Wert ohne bzw. mit Berücksichtigung des Tunings. Der angegebene Wert  $Tune^*$  gibt die detektierte Tuningabweichung des jeweiligen Songs vom Standardtuning in Halbtönen an. Beispielsweise ist der Song *Lovely Rita* laut unserer Berechnung ca. 0.67 Halbtöne höher gestimmt als in der Standardstimmung. Man kann nun aus der Tabelle ablesen, dass der Ausgleich von Tuningunterschieden einen sehr grossen Einfluss auf das Ergebnis der Akkorderkennung hat. Beispielsweise wurden für den Song *Lovely Rita* ohne Berücksichtigung des Tunings nur 3% der Akkorde richtig erkannt, während nach Berücksichtigung des Tunings über 64.2% der Akkorde richtig erkannt wurden (siehe Abb. 2). Im Schnitt hat die Berücksichtigung des Tunings für die ausgewählten Songs in der Tabelle einen Anstieg des F-Measure-Wertes von 30% auf 53.1% zur Folge, insgesamt für alle 180 Songs einen Anstieg des F-Measure-Wertes um ca. 3%.

## Zusammenfassung

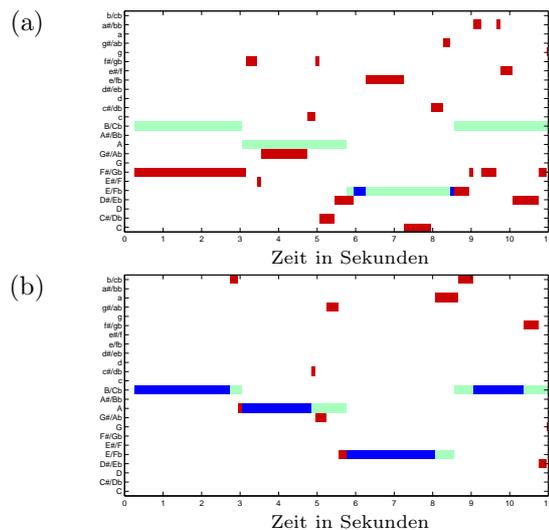
Unsere Experimente verdeutlichen verschiedene Probleme, die bei der Akkorderkennung auftreten. Hierbei hat sich insbesondere gezeigt, dass der Ausgleich von Tuningabweichungen für die Akkorderkennung eine zentrale Rolle spielt. Das Tuning sollte daher in jedem Fall bei der Akkorderkennung mitberücksichtigt werden.

Song	Tune	$F_{orig}$	$F_{tune}$
Lovely Rita	0.67	0.030	0.642
Strawberry Fields Forever	0.25	0.519	0.547
Wild Honey Pie	0.5	0.078	0.366
Ticket To Ride	0.25	0.361	0.604
Another Girl	0.67	0.181	0.446
Boys	0.33	0.304	0.434
You've Got To Hide Your Love Away	0.33	0.577	0.704
Do You Want To Know A Secret	0.5	0.346	0.503
Durchschnitt		0.300	0.531
Durchschnitt (180 Songs)		0.526	0.559

**Tabelle 1:** Vergleich der Ergebnisse der Akkorderkennung mit und ohne Ausgleich von Tuningabweichungen



**Abbildung 1:** Akkorderkennung für *Let it be*, T. 1-4.



**Abbildung 2:** Akkorderkennung für den Anfang von *Lovely Rita*. (a) Ohne Tuningausgleich, (b) Mit Tuningausgleich.

## Literatur

- [1] Alexander Sheh and Daniel P. W. Ellis. Chord segmentation and recognition using EM-trained hidden markov models. In *Proc. ISMIR, Baltimore, Maryland, USA, 2003*.
- [2] Christopher Harte, Mark Sandler, and Samer Abdallah. Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proc. ISMIR, London, UK, 2005*.
- [3] Kyogu Lee. Identifying cover songs from audio using harmonic representation. In *extended abstract submitted to Music Information Retrieval eXchange task, Victoria, BC, Canada, 2006*.
- [4] Meinard Müller. *Information Retrieval for Music and Motion*. Springer, 2007.