

Alexander Lerch [zplane.development, Berlin]

Ein Ansatz zur automatischen Erkennung der Tonart in Musikdateien

An Approach to Automatic Key Detection of Music Files

Es wird ein Verfahren zur automatischen Erkennung der Tonart von Musikdateien vorgestellt. Das Verfahren analysiert mittels einer Filterbank den Tonvorrat des Eingangssignals, der in einem Tonvektor zusammenfasst wird. Dabei sind sowohl mehrstimmige als auch einstimmige Eingangssignale zulässig. Mit Hilfe eines Nearest-Neighbour-Classifiers wird anschließend das wahrscheinlichste Ergebnis für den extrahierten Tonvektor bestimmt. Parallel zur Analyse des Tonvorrats wird die Stimmhöhe des Kammertons detektiert, um eine gleichbleibende Erkennungsrate für Signale unterschiedlicher Stimmhöhe zu gewährleisten.

1 EINFÜHRUNG

Die rapide zunehmende Zahl von Audiosignalen, die beispielsweise im Internet Millionen von Nutzern zur Verfügung stehen, erfordert neue Ansätze zur Katalogisierung und Indizierung dieser Daten. Die zur Zeit verfügbaren Metadaten bzw. Informationen über die eigentlichen Inhalte werden, so sie überhaupt existieren, manuell erstellt. Eine gezielte Suche nach Musikstücken mit bestimmten Eigenschaften wie beispielsweise Genre, Instrumentierung, Tempo, Rhythmus oder Tonart ist aus diesem Grund zur Zeit nicht verfügbar. Um eine solche intuitive und aus Anwendersicht naheliegende Suche in einer großen Datenmenge zu ermöglichen, sind daher Ansätze zur automatischen Extraktion solcher Metadaten dringend erforderlich. Diese Arbeit behandelt mit der Präsentation eines Verfahrens zur automatischen Erkennung der Tonart einen Teilbereich dieses weitläufigen Forschungsgebietes.

Neben der automatisierten Generierung von Metadaten für die Katalogisierung bspw. für die Suche im Internet finden sich mögliche Anwendungen einer automatischen Tonarterkennung einerseits in Applikationen für loop-basierte Musikproduktion, wo mehrere Ausschnitte oder Loops miteinander gemischt werden; andererseits kann eine Tonarterkennung auch zur automatischen Generierung von Playlists verwendet werden, wo die Tonartbeziehung zweier Stücke ebenfalls zu einer Aussage über die „Kompatibilität“ der Einzelstücke beitragen kann. Weiterhin kann die Vorgabe der Tonart Algorithmen wie beispielsweise eine

Tonhöhenenerkennung (Pitch-Tracking) verbessern, da innerhalb eines gegebenen tonalen Kontextes nicht mehr alle Töne der Halbtonskala als gleich-wahrscheinlich angenommen werden müssen.

1.1 Definitionen

Die *Tonart* eines tonalen Musikstückes lässt sich durch ein bestimmtes Tongeschlecht, bezogen auf eine Transpositionsstufe beschreiben. Die in dieser Arbeit betrachteten Tongeschlechter sind Dur und Moll, als Transpositionsstufen gelten die 12 Tonklassen, welche die 12 Töne der westlichen Tonskala in temperierter Stimmung unter Vernachlässigung der Oktavlage darstellen. Insgesamt werden also 24 mögliche Tonarten betrachtet. Der Tonvorrat einer Tonart setzt sich zusammen aus den Tönen einer dem Tongeschlecht entsprechenden diatonischen Tonleiter, beginnend bei der jeweiligen Transpositionsstufe. Obwohl die meisten Musikstücke eine bestimmte und identifizierbare Grundtonart besitzen, kann die Tonart über den zeitlichen Verlauf des Stückes einmal oder mehrmals variieren (Modulation). Modulationen sind i.a. wahrscheinlicher hin zu verwandten Tonarten, wobei der „Verwandtschaftsgrad“ zweier Tonarten sich aus ihrer Entfernung im Quintenzirkel abschätzen lässt.

Der Kammerton ist der Ton, auf den eine Gruppe von Instrumenten gestimmt wird. Er entspricht der klingenden Tonhöhe des eingestrichenen *a*. Obwohl seine Frequenz mit 440Hz oft als fest angenommen wird, kann die tatsächliche Frequenz des Kammertons in der Praxis aus unterschiedlichen Gründen (z.B. Verwendung alter Instrumente, persönliche Klangvorlieben der Musiker) von 440Hz durchaus abweichen. Die Frequenz des Kammertons wird im folgenden als *Stimmhöhe* bezeichnet.

1.2 Übersicht

Die Struktur dieser Ausarbeitung ist wie folgt: Nach der Beschreibung der Anforderungen und Restriktionen des Algorithmus und einer exemplarischen Literaturübersicht werden die einzelnen Verfahrensschritte Frequenztransformation, Erstellung des Tonklassenvektors, Schätzung der Tonart aus dem Tonklassenvektor und die Schätzung der Stimmhöhe erläutert. Der darauf folgende Abschnitt zur Evaluierung beschreibt die verwendete Testdatenbank und wertet die Ergebnisse aus. In der abschließenden Schlussbemerkung wird ein kurzer Ausblick auf mögliche Erweiterungen und Optimierungen des Verfahrens gegeben.

2 ANFORDERUNGEN

Das Ziel bei der Entwicklung des vorgestellten Verfahrens war die Klassifizierung einer Musikdatei in eine von 24 Tonartklassen, bestehend aus jeweils Dur- und Moll-Tonarten für jede Transpositionsstufe. Weitere Anforderungen waren die Robustheit des Verfahrens gegenüber Instrumentierung, Klangfarbe und Aussteuerung des Eingangssignals und die Verarbeitung sowohl ein- als auch mehrstimmiger Signale. Es werden nur Eingangssignale gleichbleibender Tonart, d.h. ohne Modulationen, zugelassen, was den Einsatzbereich der Tonarterkennung hauptsächlich auf populäre Musik einschränkt, in der Modulationen relativ selten auftreten.

Die Stimmhöhe unterschiedlicher Aufnahmen wird im allgemeinen nicht übereinstimmen, so dass weiterhin die automatische Adaption an die Stimmhöhe des Eingangssignals gefordert werden muss.

Das Verfahren soll im Hinblick auf die erforderliche Rechenleistung möglichst einfach gehalten werden, um es sinnvoll in bestehende Softwareapplikationen integrieren zu können.

3 AUTOMATISCHE TONARTERKENNUNG

In der Vergangenheit wurden verschiedene Verfahren zur Schätzung der Tonart vorgestellt, die in den meisten Fällen keine Audiodaten verarbeiten, sondern Informationen basierend auf dem Notenbild oder einer MIDI-Datei verwenden. An dieser Stelle soll lediglich ein kurzer Überblick über die auf Krumhansls Verfahren [Kru90] basierenden Veröffentlichungen gegeben werden. Auch Krumhansl verwendet keine Audiodaten zur Schätzung der Tonart, sondern direkt aus dem Notenbild extrahierbare symbolische Informationen und schätzt die Tonart durch die für jede Tonart charakteristische Verteilung der Häufigkeit der 12 Tonklassen (detaillierte Beschreibung s.u.).

In jüngster Zeit gibt es Ansätze, die Tonart eines Stückes oder von Ausschnitten daraus direkt aus den Audiodaten zu schätzen. Diese basieren zumeist auf der Arbeit von Krumhansl. Zur Erzeugung des Tonklassenvektors werden hier unterschiedliche Ansätze gewählt: während Purwins, Blankertz und Obermayer die FFT-ähnliche Constant-Q-Transformation verwenden [PBO00], wird von Cremer und Derboven eine Frequency-Warped FFT angewendet [CD04]. Pickens, Bello et al. verwenden im Gegensatz zu Krumhansl keinen Tonklassenvektor, sondern einen Akkordklassenvektor, der mit Hilfe eines automatischen Transkriptionssystems erzeugt wird [PBM+02].

Das einzige dem Autor bekannte Verfahren zur Schätzung der Frequenz des Kammertons wurde von Scheirer 1995 [Sch95] beschrieben. Hierbei wird das Audiosignal mit einem Set von Bandpassfiltern gefiltert, die temperiert bezüglich des Kammertons gestimmt sind. Die Filterung wird mehrmals mit jeweils leicht veränderter Stimmhöhe durchgeführt; anschließend wird die Frequenz des Kammertons des Durchlaufs mit der höchsten mittleren Energie an den Filterausgängen als die gesuchte Frequenz gewählt. Scheirer wählt die Tonlagen der Einzelfilter allerdings ausgehend vom Notenbild (genauer: einer MIDI-Datei), ein Ansatz, der für ein System ohne Noteninformation nicht anwendbar ist.

4 VERFAHREN

Das hier vorgestellte Verfahren lässt sich in zwei grundsätzliche Schritte aufteilen: die Extraktion eines repräsentativen Tonklassenvektors (im folgenden *pitch chroma*, vgl. z.B. Bartsch und Wakefield [BW01]) für das Eingangssignal und die Schätzung der wahrscheinlichsten Tonart aufgrund des berechneten pitch chromas.

Das pitch chroma ist ein Vektor, dessen 12 Elemente für die 12 Töne der Tonleiter ohne Berücksichtigung der Oktavlage stehen. Das Verhältnis der Elemente zueinander ist bestimmt durch die Häufigkeit und die Energie, mit der diese Töne unabhängig von ihrer Oktavlage innerhalb der Audiodatei auftreten.

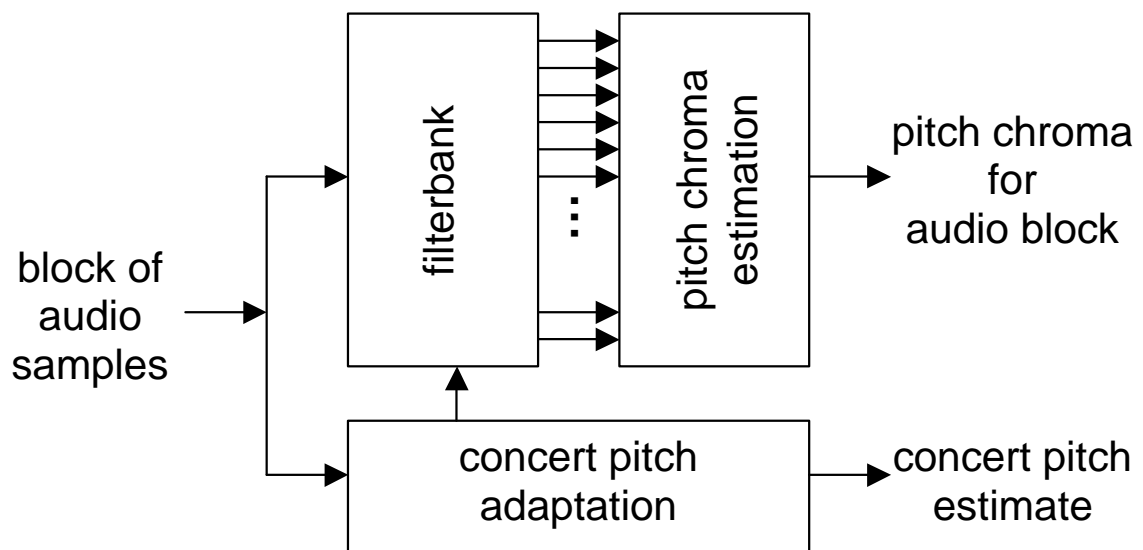


Abbildung 1: Berechnung des pitch chromas für einen Block von Audiosamples

Aus der Verteilung der Einzelelemente des pitch chromas des Gesamtsignals wird anschließend die Tonart mit der höchsten Wahrscheinlichkeit bestimmt. Abbildung 1 zeigt

das Blockschaltbild zur Berechnung des pitch chromas für einen Audioblock. Die einzelnen Teilschritte werden im folgenden erläutert.

4.1 Filterbank

Aufgrund der geforderten Unabhängigkeit des Verfahrens gegenüber Instrumentierung und Klangfarbe wird bei der Erstellung des pitch chromas von Modellen z.B. zu Intensität und Verlauf der Obertöne abgesehen. In dem vorliegenden Verfahren wird das Signal mittels einer Filterbank frequenztransformiert. Abbildung 2 zeigt den Frequenzgang der verwendeten Filterbank. Die Untersuchung alternativer Frequenztransformationen wie z.B. der Constant-Q-Transformation haben keine signifikanten Unterschiede im Ergebnis gezeigt. Die Zahl der Filter ist die Zahl von Halbtönen pro Oktave (zwölf), multipliziert mit der Anzahl der Oktaven, die analysiert werden sollen. Die Güte der Einzelfilter ist konstant, und die Abstände der Mittenfrequenzen der Filter entsprechen einem Halbtonschritt der temperierten Stimmung. Die Verwendung einer anderen Stimmung ist problematisch, da die einzelnen Tonabstände abhängig von der verwendeten Tonart bestimmt werden müssten, die zu diesem Zeitpunkt noch nicht bekannt ist. Der Ausgang jedes Filters ist die mittlere Energie des jeweiligen Frequenzbandes innerhalb des Analysefensters der Länge 40ms.

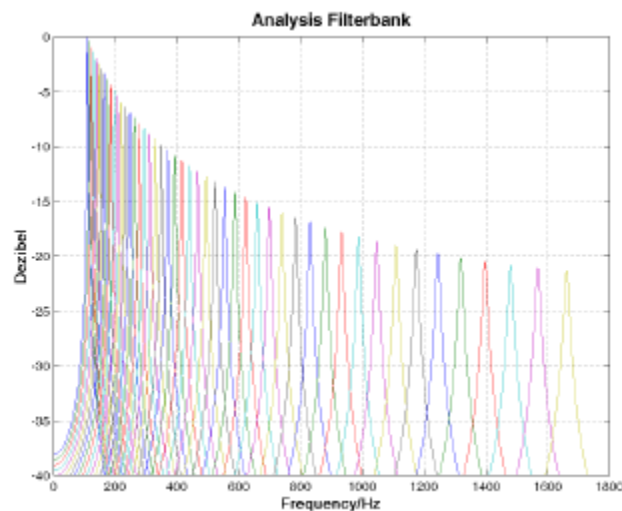


Abbildung 2: Darstellung der Betragsfrequenzgänge der Einzelfilter der nicht normierten Filterbank

4.2 Bestimmung des Pitch Chromas

Das pitch chroma ist eine Darstellung der Tonhöhe ohne Oktavinformation, d.h. eine Darstellung von 12 Tonklassen, denen die Namen *A*, *Ais*, *H*, *C*, ..., *Gis* zugeordnet werden können. Das pitch chroma eines einzelnen Analysefensters berechnet sich aus den Filterausgängen über die Summierung über alle Oktaven. Das pitch chroma der Gesamtaudiodatei wird über die Mittelung der pitch chromas aller Analysefenster bestimmt. Da die Filterbank temperiert gestimmt ist, sind enharmonische Verwechslungen möglich, d.h.

die Tonklasse *Dis* entspricht beispielsweise der Tonklasse *Es*. Abbildung 4 zeigt ein typisches pitch chroma der Tonart *D-Dur*.

4.3 Schätzung der Tonart

Es ist naheliegend, dass die Relationen der einzelnen Tonklassen des pitch chromas über die Tonart des Stückes Aufschluss geben. So ist zu erwarten, dass tonartfremde Tonklassen sehr geringe Werte im Vergleich z.B. zu Tonika oder Dominante haben. Es bleibt jedoch die Frage, wie die Referenzvektoren für die typische Verteilung bestimmt werden können. Die sogenannten Probe Tone Ratings wurden von Krumhansl [Kru90] ermittelt, indem einer Gruppe von Testhörern ein beliebiger Ton vorgespielt wurde, nachdem der tonale Kontext durch eine kadenzartige Akkordfolge etabliert war. Die Aufgabe der Testhörer war nun, auf einer Skala von 1 bis 7 anzugeben, wie gut der gespielte Ton in den gegebenen Kontext passt. Abbildung 3 zeigt die Resultate des Experiments. Wie zu erwarten war, finden sich die höchsten („passendsten“) Werte bei dem Grundton der jeweiligen Tonart. Auch die Quinte und Terz zeigen sowohl in Dur als auch Moll deutliche Maxima.

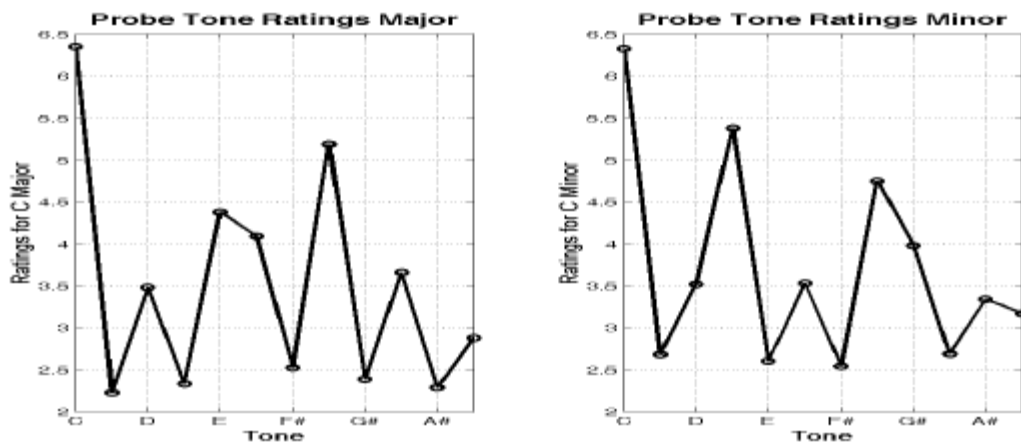


Abbildung 3: Probe Tone Ratings nach Krumhansl für Dur (links) und Moll (rechts), aufgetragen für C-Dur bzw. c-moll

Krumhansl fand eine bemerkenswert hohe Korrelation zwischen den Probe Tone Ratings und einer Messung der Häufigkeit bzw. der Tondauern der entsprechenden Töne in verschiedenen Musikstücken und nutzt diese Ergebnisse für ein Verfahren zur Tonarterkennung, auf dem der hier vorgestellte Algorithmus basiert.

Bei der in dieser Arbeit verwendeten Berechnung des pitch chromas werden symbolische Informationen wie Tonhäufigkeit und –dauer nicht oder nur implizit berücksichtigt, während

als weiterer Parameter die Energie des Tons einfließt. Dies ist aber nicht zwangsläufig ein Nachteil, wie die Ergebnisse zeigen. Abbildung 4 stellt beispielhaft das gemessene pitch chroma eines Popstückes in D-Dur neben den Probe Tone Ratings dar (beide normiert).

Die Probe Tone Ratings für Dur und Moll werden als Referenzvektoren mit jeweils 12 Varianten, entsprechend den möglichen 24 Tonartklassen, gespeichert. Die wahrscheinlichste Tonart ist dann die, deren Referenzvektor den geringsten Abstand (Nearest Neighbour) zum berechneten pitch chroma aufweist. Als Abstandsmaß wird die euklidische Distanz gewählt. Die Schätzung der Tonart über das euklidische Abstandsmaß hat sich für das vorgestellte System als robuster erwiesen als die Korrelation der Vektoren wie sie z.B. bei Krumhansl [Kru90] und Cremer [CD04] verwendet wird.

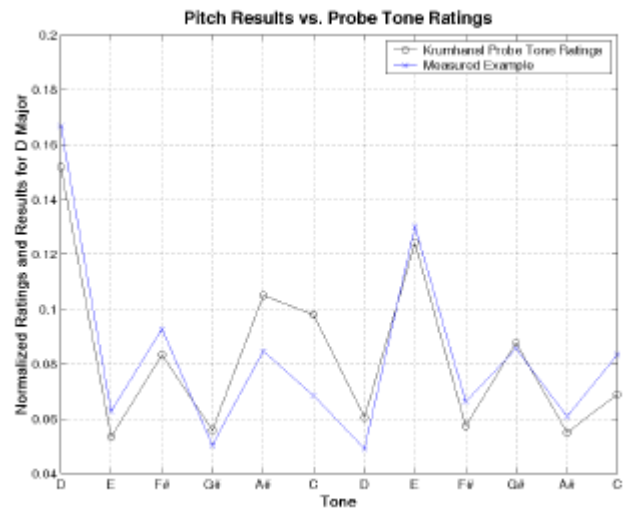


Abbildung 4: Vergleich von Krumhansls Probe Tone Ratings mit einem gemessenen pitch chroma

4.4 Erkennung der Höhe des Kammertons

Wären die Mittenfrequenzen der Einzelfilter der Filterbank fest (z.B. bezüglich des Kammertons 440Hz), so wären die Erkennungsraten bei Aufnahmen anderer Stimmhöhe wesentlich schlechter, da die in dem Signal auftauchenden Grundfrequenzen der Einzeltöne nicht mehr den Mittenfrequenzen der Filter entsprechen. Um dieses Problem zu umgehen, werden die Mittenfrequenzen pro Analysefenster neu gesetzt. Dazu wird das Audiosignal wiederum mittels einer Filterbank transformiert, deren Einzelfilter sehr steilflankig sind. Abbildung 5 stellt exemplarisch den Betrags-

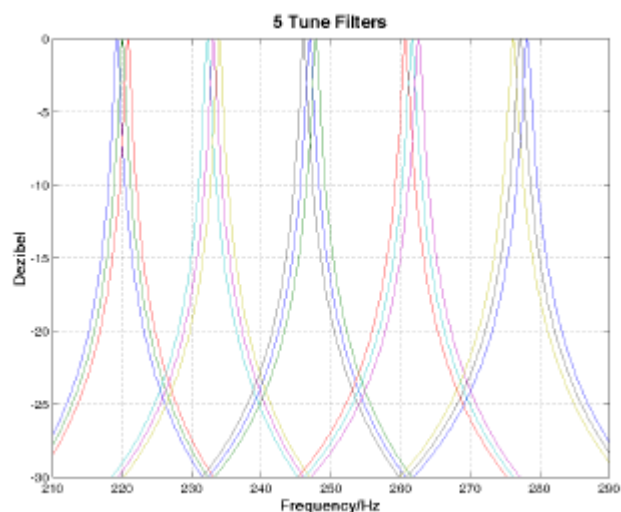


Abbildung 5: Betragsfrequenzgang der Filter zur Adaption der Stimmhöhe für fünf Halbtöne

frequenzgang für fünf Halbtöne dar. Jeder Halbton wird hier nicht nur durch ein Filter bestimmt, sondern jeweils durch drei symmetrisch zur Mittenfrequenz angeordnete Filter. Die Ausgänge der Filter werden über alle Halbtöne gemittelt, so dass für jedes Analysefenster ein Ergebnisvektor mit drei Elementen t_1 , t_2 , t_3 resultiert. Das mittlere Element t_2 entspricht dabei der Energie des Ausgangs des mittleren Filters, die anderen beiden die mittlere Energie bei leicht höher oder niedriger liegenden Frequenzen. Die Adaptionregel zur Berechnung der Stimmfrequenz f_F der Filterbank für das nächste Analysefenster n berechnet sich dann mit

$$f_F(n) = \mathbf{h} \cdot \frac{t_3 - t_1}{t_1 + t_2 + t_3} \cdot f_F(n-1)$$

wobei der Faktor \mathbf{h} die Adaptionsgeschwindigkeit bestimmt. Die ständige Adaption des Filters hat auch zur Folge, dass Dateien mit variierender, z.B. langsam abfallender Stimmhöhe korrekt verfolgt werden können. Der Initialisierungswert der Stimmfrequenz ist 440Hz.

5 EVALUIERUNG

Die Testdatenbank zur Evaluierung der Tonarterkennung besteht aus insgesamt 210 Audiodateien mit einer Gesamtlänge von mehr als 15 Stunden mit einem korrespondierenden Datenvolumen von ca. neun Gigabyte. Die Datenbank beinhaltet sowohl Signale in CD-Qualität als auch decodierte MP3s unterschiedlicher Qualität und ist aus unterschiedlichen Genres (wie z.B. Folk, Jazz, Pop, Rock) und dementsprechend unterschiedlichen Instrumentierungen zusammengesetzt. Die Testdateien wurden willkürlich aus einer umfassenderen Datenbank gewählt, allerdings mit der Restriktion, dass die Tonart sich im Verlauf nicht ändert. Die Tonarten der Stücke, die bei der Evaluierung als Referenz gelten, wurden größtenteils aus Internetdatenbanken entnommen und „manuell“ mit der Stimmgabel verifiziert.

5.1 Test 1: Populärmusik

Die zufällige Erkennungsrate, d.h. die Wahrscheinlichkeit einer korrekten Tonarterkennung im Falle der zufälligen Zuweisung eines Ergebnisses liegt für die 24 Tonartklassen bei $1/24 = 4.2\%$. Die Rate der korrekten Klassifizierungen des vorgestellten Verfahrens liegt mit der o.g. Datenbank bei 77%. Eine Analyse der falsch klassifizierten 23% liefert folgendes Ergebnis: mehr als zwei Drittel der Fehlerkennungen sind direkt verwandte Tonarten (Paralleltonart, Tonart der Subdominante, Tonart der Dominante). Diese Fehlklassifizierungen sind daher für viele Applikationen erträglicher als Fehlklassifizierungen zu im Quintenzirkel weiter entfernten Tonarten. Tabelle 1 zeigt in der Zeile Populärmusik die detailliert

aufgeschlüsselten Ergebnisse, die zusätzlich in Abbildung 6 graphisch dargestellt sind. Die zusätzliche Spalte Dur-Moll-Verwechslung zeigt die Häufigkeit von Verwechslungen zwischen Dur- und Molltonarten gleichen Grundtons.

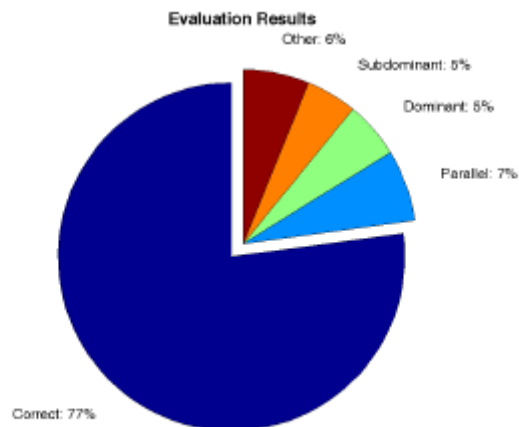


Abbildung 6: Darstellung des Verhältnisses von korrekten und falschen Ergebnissen (Test 1)

5.2 Test 2: Klassik (1)

Obwohl der Ansatz des Verfahrens auf der Annahme einer unveränderlichen Tonart beruht, wird in einem zweiten Test untersucht, wie die Tonarterkennung auf Eingangssignale mit Modulationen reagiert. Dazu wird das Verfahren mit einer kleineren Testdatenbank von insgesamt 70 Audiodateien und einem entsprechenden Datenvolumen von 2.4 Gigabyte evaluiert. Die Testsignale dieser Datenbank beinhalten ausschließlich klassische Musik der Epochen Barock bis Romantik und variieren in der Besetzung von Klavier Solo über Streichquartett, Orchester bis zu Ausschnitten aus Oratorien. Die Erkennungsrate liegt in diesem Fall, wie aufgrund der Signale mit wechselnder Tonart zu erwarten war, deutlich niedriger: lediglich 58.6% der Testdateien konnten korrekte Tonarten zugeordnet werden. Die Auswertung der Fehlklassifizierungen zeigt, dass ca. 25% der Dateien fälschlicherweise der Tonart der Quinte (Dominante) zugeordnet wurden. Die aufgeschlüsselten Ergebnisse sind in Tabelle 1 in der Zeile *Klassik (1)* einzusehen.

5.3 Test 3: Klassik (2)

Mit einer drastischen Vereinfachung konnte das Ergebnis für die Klassik-Testdatenbank deutlich verbessert werden: unter der Annahme, dass ein Großteil der Stücke in der Grundtonart endet, werden nur die jeweils 20 letzten Sekunden der Testdateien analysiert. Die Erkennungsrate konnte in diesem Fall signifikant auf 72.9% erhöht werden (s. Tabelle 1,

Zeile *Klassik (2)*). Es bedarf Untersuchungen mit einer größeren Datenbank zur Verifizierung dieser vorläufigen Ergebnisse.

	Korrekte Erkennung	Fehlerkennungen				
		Paralleltonart	Dominante	Subdominante	Dur/Moll-Verwechslung	Sonstige Fehler
Popularmusik	77.1%	6.7%	5.2%	4.8%	2.9%	3.3%
Klassik (1)	58.6%	7.1%	25.7%	1.4%	0.0%	7.1%
Klassik (2)	72.9%	0.0%	11.4%	4.3%	4.3%	7.1%

Tabelle 1: korrekte Erkennungsrate (auf eine Dezimalstelle gerundet) und Analyse der falsch klassifizierten Tonarten für alle drei Testläufe

5.4 Test 4: Stimmhöhe

Eine Methodik zur objektiven Evaluierung der Stimmhöhenerkennung existiert nicht, da für die Dateien aus der Testdatenbank keine Referenzdaten vorliegen und eine nachträgliche Extrahierung kaum möglich bzw. anzweifelbar ist. Daher wurde das Verfahren mit künstlichen Signalen getestet. Die maximalen Abweichungen von der tatsächlichen Stimmhöhe liegen hier in der Größenordnung von ca. 4cent. Im Falle von natürlichen Musiksignalen ist anzunehmen, dass die maximale Abweichung etwas höher ist. Die Aussagekraft der Ergebnisse konnte nur mit Gehör durch Addition eines Sinussignals zum Testsignal für einen stichprobenhaften Ausschnitt der Testdateien subjektiv verifiziert werden.

Abbildung 7 zeigt, wie die gemessene Stimmhöhe für die Testdateien beider Datenbanken von der Frequenz 440Hz abweicht. Hierbei fällt ein Trend zu höheren Kammertönen als 440Hz auf, der sich größtenteils auf den Einfluss der klassischen Testdateien auf das Gesamtergebnis zurückführen lässt.

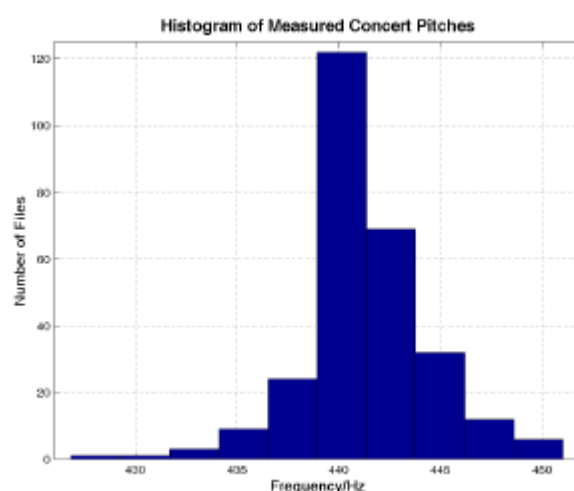


Abbildung 7: Verteilung der Stimmhöhe der Testdateien in Klassen der Breite 2Hz um die Frequenz 440Hz

5.5 Test 5: Rechenleistung

Die vorliegende –noch nicht optimierte– Implementierung des Verfahrens benötigt, bezogen auf einen Prozessor der Taktfrequenz 1GHz, ungefähr ein 11-tel der Dateilänge zur Berechnung des Ergebnisses. Dieser Wert wurde über die (größere) Popularmusik-Datenbank gemessen.

6 SCHLUSSBEMERKUNG UND AUSBLICK

Es wurde ein effizientes Verfahren zur automatischen Erkennung der Tonart von Musikdateien vorgestellt, das mit einer Erkennungsrate von ca. 77% brauchbare Ergebnisse liefert. Können auch Fehlentscheidungen hin zu verwandten Tonarten toleriert werden, so beträgt die Wahrscheinlichkeit einer tolerierbaren Entscheidung ca. 93%. Aufgrund der gering gehaltenen algorithmischen Komplexität und der daraus resultierenden geringen Rechenleistung kann der Algorithmus in unterschiedlichsten Applikationen zumindest zu einer ersten Schätzung der Tonart sinnvoll eingesetzt werden.

Das größte Potential für eine weitere Verbesserung der Erkennungsrate scheint vor allem in einer Optimierung der Klassifizierung des pitch chromas zu liegen. Als Klassifikatoren könnten beispielsweise GMMs (Gaussian Mixture Models), SVMs (Support Vector Machines) oder Neuronale Netze zu verbesserten Ergebnissen führen. Diese aufwendigeren Klassifikatoren erfordern allerdings im Vergleich zu dem hier beschriebenen einfachen Nearest-Neighbour-Klassifikator ein Training, was aber unter Umständen zu besseren, d.h. besser an die verwendete Frequenztransformation angepassten Referenzvektoren für Dur und Moll führen kann.

Weiterhin soll die Erweiterung des Verfahrens im Hinblick auf die Erkennung von Modulationen bzw. lokalen Tonarten untersucht werden. Erste Ergebnisse einer Tonarterkennung in kürzeren zeitlichen Abschnitten mittels eines Moving-Average-Filters sind allerdings nicht vielversprechend, so dass wahrscheinlich aufwendigere Ansätze zur Glättung des Zwischenergebnisses erforderlich sind (vgl. Shmulevich und Yli-Harja [SY00]).

Eine Version zur Evaluierung der Implementierung steht online unter der Adresse <http://www.zplane.de> zur Verfügung.

[Literaturhinweise]

- [BW01] Bartsch, Mark A.; Wakefield, Gregory H.: *To Catch a Chorus: Using Chroma-Based Representations for Audio Thumbnailing*, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'01), New Paltz, 2001
- [CD04] Cremer, Markus; Derboven, Claas: *A System for Harmonic Analysis of Polyphonic Music*, Proc. of the of the AES 25th International Conference: Metadata for Audio, London, 2004
- [Kru90] Krumhansl, Carol L.: *Cognitive Foundations of Musical Pitch*, Oxford University Press, New York, 1990
- [PBM+02] Pickens, Jeremy; Bello, Juan Pablo; Monti, Giuliano; Crawford, Tim; Dovey, Matthew; Sandler, Mark; Byrd, Don: *Polyphonic Score Retrieval Using Polyphonic Audio Queries: A Harmonic Modeling Approach*, Proc. of the International Conference on Music Information Retrieval (ISMIR'02), Paris, 2002
- [PBO00] Purwins, Hendrik; Blankertz, Benjamin; Obermayer, Klaus: *A new method for tracking modulations in tonal music in audio data format*, Proc. of the International Joint Conference on Neural Network (IJCNN'00), Como, 2000
- [Sch95] Scheirer, Eric. D.: *Extracting Expressive Performance Information from Recorded Music*, M.S. Thesis, Massachusetts Institute of Technology, Cambridge, 1995
- [SY00] Shmulevich, Ilya; Yli-Harja, Olli: *Localized Key-Finding: Algorithms and Applications*, Music Perception 17(4), 2000