

Anforderungen, Möglichkeiten und Grenzen der automatischen Musikanalyse

UWE BAUMANN (Fachgebiet Akustische Kommunikation, Techn. Univ. München)

Einleitung

Die Möglichkeit einer objektiven Analyse von Musik findet sowohl bei Musikwissenschaftlern als auch bei praktizierenden Musikern breites Interesse. Forscher aller Epochen bearbeiteten diese Fragestellung mit wissenschaftlichen Methoden; stellvertretend seien hier Pythagoras, Ohm, Helmholtz und Planck genannt. Dennoch ist die bisher aus objektiven Messungen zusammengetragene Datenmenge vergleichsweise klein. Ursache hierfür ist die im allgemeinen hohe Komplexität von Musik, verbunden mit einem Mangel an Werkzeugen zur Reduktion dieser Komplexität auf die für die menschliche Empfindung relevanten Parameter. Bei den bisher durchgeführten Analysen beschränkte man sich aus diesem Grunde auf wenige Parameter wie zum Beispiel das Zeitintervall zwischen einzelnen Tönen oder die Grundtonhöhe. Trotz der Beschränkung auf wenige Parameter ist der Arbeitsaufwand zur Gewinnung von objektiven Messdaten bei Verwendung der bisher überwiegend manuellen Analysemethoden ausgesprochen hoch. Repp [6] untersuchte 28 Interpretationen der „Träumerei“ von F. Schubert und mußte für jede dieser Aufnahmen zur Feststellung der Einsatzzeitpunkte ca. 200 charakteristische Stellen in einer Wellenformdarstellung des digitalisierten Schallsignales an einem Computermonitor markieren. Der Zeitaufwand und die Fehleranfälligkeit dieses Verfahrens sind entsprechend hoch; dabei kam es zu Ungenauigkeiten bis 20 Prozent. Loosen [5] untersuchte die Intonation von acht professionellen Violinisten, die eine drei Oktaven umfassende C-Dur Tonleiter spielen und dabei auf jedem Ton ohne Vibratoanwendung drei Sekunden ausharren mußten. Als Maß für die Intonation wählte Loosen die Frequenz der ersten Harmonischen, welche er durch Tiefpass-Filterung und nachfolgender gemittelter Frequenzzählung der Tonbandaufnahmen gewann.

Diese wenigen Beispiele zeigen bereits, daß eine automatische Datenaufbereitung von Musiksignalen sinnvoll ist; einerseits lassen sich erhebliche Arbeitserleichterungen bei der Gewinnung objektiver Daten erzielen, andererseits werden Ablesefehler vermieden.

Anforderungen an eine automatische Musikanalyse und realisierte Ansätze

Tab.1 listet neben einigen Grundanforderungen auch weitergehende Leistungen auf. Besondere Aufmerksamkeit neben den aufgeführten Punkten erfordert die Behandlung mehrstimmiger Musik; dazu ist ein Verfahren zur Extraktion der Einzelstimmen erforderlich.

Ein Ansatz zur automatischen Musikanalyse wird z.B. in [1] beschrieben. Das *Visa Project* benötigt eine Fülle von Benutzereingaben, insbesondere die Segmentierung und die Bestimmung musikalischer Intervalle mußten von Hand durchgeführt werden. Ein diskret aufgebauter Grundfrequenzdetektor liefert unter Beeinflussung des Quellsignales durch Raumhall fehlerhafte Tonhöhen. Es ist nur für einstimmige Musik verwendbar.

Ein weiterer Ansatz findet sich in [4]. Die dort beschriebene *Sentiment Extraction in Music* erlaubt eine automatische Transkription einfacher polyphoner Klaviermusik mit recht guter Erkennungsrate. Die Tonhöhenerkennung erfolgt unter Benutzung einer „Frequenzkarte“; zur korrekten Erkennung muß die 1. Harmonische vorhanden sein. Es werden nur Tonstufen in einem Halbtonraster zugelassen; eine Auswertung der Tonhöhenfeinstruktur wird nicht durchgeführt. Kürzere Notendauern als Viertelnoten werden nicht erkannt. Die Takterkennung wird durch eine Differenzierung des Signalpegels verbunden mit der Bestimmung von Betonungen ermöglicht. Eine Melodielinie wird nach heuristischen Regeln

Grundanforderungen	Erweiterte Anforderungen	Spezielle Untersuchungen
Tonhöhenbestimmung	Erkennung von Tonhöhenintervallen Analyse der benutzten Intonation (Stimmung)	Tonhöhenfeinstruktur (Vibrato, Sweep) Akkordsymbolzuordnung
Dauer von Tönen und Pausen Erkennung von Betonungen	Rhythmus- und Takterkennung Erkennung spezifischer Änderungen der Einsatzzeitpunkte	Bestimmung des Notenwertes Bestimmung der Tempofeinstruktur Erkennung musikalischer Phrasen
Erkennung spektraler Charakteristika	Zuordnung von Klangfarben zu Instrumenten	Qualitätsbeurteilung des Instrumentes

Tab. 1: Anforderungen an eine automatische Musikanalyse

ermittelt.

Ein Verfahren zur *Stimmentrennung in polyphoner Musik* [2] erlaubt die Trennung zweistimmig polyphoner Musik. Am Beispiel einer Invention von J. S. Bach wurde die Rekonstruktion der Einzelstimmen aus dem Signalgemisch demonstriert.

Möglichkeiten

Das in [2] beschriebene Verfahren ist nur für synthetische Klänge mit synchronem Einsatz der zu einem Einzelton gehörenden Harmonischen geeignet. Um auch natürliche Klänge verarbeiten zu können, wurde ein zeitliches Toleranzfenster für die Gruppierung von Teiltonlinien zu einem Cluster eingeführt. Zusätzlich wurde das Verfahren durch eine Interpolationsmöglichkeit bei der Bildung von Tonhöhenkonturlinien erweitert und erlaubt nun die Bearbeitung einfach strukturierter natürlicher Musiksignale. Allerdings ist immer noch ein Benutzereingriff erforderlich: Die aus der Analyse gewonnenen Einsatzcluster müssen noch von Hand den Einzelstimmen zugeordnet werden, da eine automatische Auswertung der spektral-zeitlichen Feinstruktur zur Zuordnung eines Clusters zu einer Instrumentenstimme noch nicht erfolgt.

Abb.2 zeigt das Teiltonzeitmuster [3] (TTZM) eines Ausschnittes aus einem einfach strukturierten Musikstück bei dem Tuba, Posaune, 2 Klarinetten und Saxophon als Begleit- und eine Trompete als Melodieinstrument zusammen spielen (Notentext in Abb.1); man erkennt die starke Überlagerung und Vermischung der Einzelinstrumente.

Die Einsatzerkennung und -gruppierung liefert z.B. für die Trompetenstimme das in Abb.2 rechts dargestellte Ergebnis. Der Melodieverlauf ist gut zu erkennen; bei $t = 2,5$ s und $t = 7,0$ s kommt es durch Saxophoneinsätze zu Störungen der Struktur höherer Harmonischer. Eine Resynthese der getrennten Einzelstimmen erlaubt eine akustische Überprüfung der Stimmenzuordnung. Die Weiterverarbeitung der Einzelstimmen kann Aufschlüsse über Intonation, Intervallstufen, Inter-Onset-Intervalle und dem Einsatz von Vibrato und Tremolo geben. Als Beispiel für die Intonations-Bestimmung dient Abb.1 rechts. Die extrahierte Trompetenstimme wurde einer geglätteten Leistungssummiierung unterzogen und über einem Frequenzbereich zwischen der Tonhöhe F4 und F5 aufgetragen. Die fünf Ton-

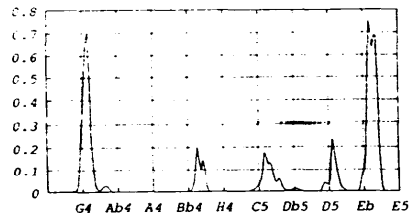


Abb. 1: Links: Notentextauszug des analysierten Musikstückes „Indien“ von H. J. Buchner (CD „Höhlenmalerei“). Rechts: Aus Abb.2 re. gewonnenes geglättetes Leistungshistogramm der extrahierten Trompetenstimme.

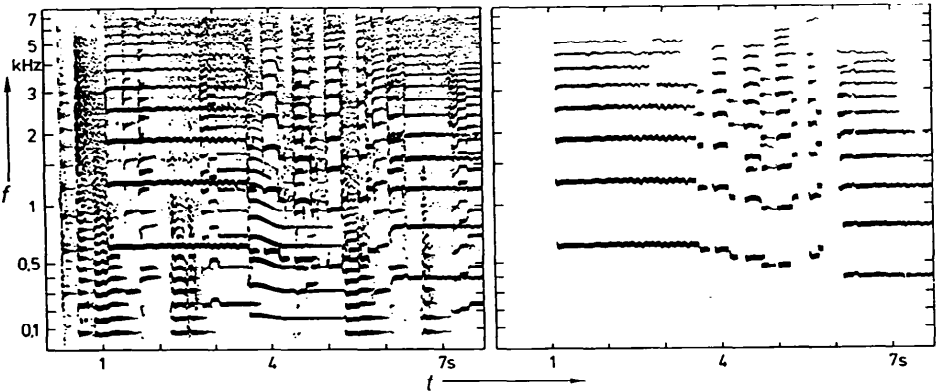


Abb. 2: Links: Teilzeitmuster „Indien“ (Notentext siehe Abb.1). Melodiestrument Trompete; Tuba, Posaune, zwei Klarinetten und Saxophon Begleitung. Rechts: Die aus dem Gesamtsignal herausgelöste Trompetenstimme.

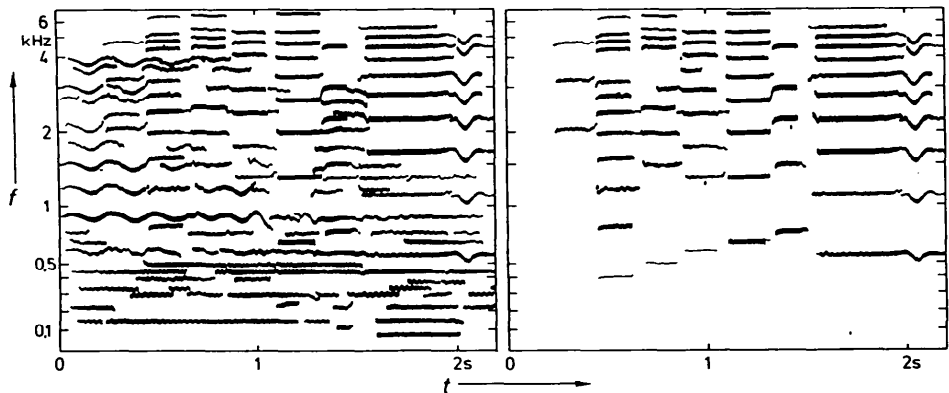


Abb. 3: Links: Teilzeitmuster eines Ausschnittes aus dem Musikstück „Just the way you are“ von B. Joel (CD „The Stranger“). Ausklingender Gesang (Vokal /a/), Beginn des Saxophonsolos (aufsteigende Linien). Rechts: Teilzeitmuster der extrahierten Saxophonstimme. Dünn: virtueller Tonhöhenverlauf der abgebildeten Stimme.

stufen der Melodie bilden sich deutlich ab, eine leichte Frequenzverschiebung der Trompetenmelodie um +10 cent bezogen auf ein A4 von 440 Hz ist ersichtlich.

Grenzen

Der unter den Grundanforderungen in Tab.1 als erstes aufgeführte Punkt „Tonhöhenbestimmung“ erscheint auf den ersten Blick trivial. Bei näherer Betrachtung, insbesondere bei komplexer, aus mehreren Stimmen zusammengesetzter Musik, ist die Leistung des menschlichen Gehörs, aus aus der Vermischung zahlreicher spektraler Komponenten mit unterschiedlichen Grundfrequenzen Tonhöhenindrücke und Melodieverläufe zu erkennen, bemerkenswert gut.

Abb.3 zeigt ein TTZM, welches die momentanen Grenzen des Verfahrens aufzeigt. Es handelt sich um einen kurzen Ausschnitt aus einem Musikstück, bei dem Gesang, E-piano und Saxophon zusammen erklingen. Die Anwendung des Trennungsverfahrens zur Gewinnung der Saxophonstimme führt zu dem in Abb.3 rechts dargestellten Ergebnis. Auffällig ist das Fehlen der tieferen Harmonischen während der ersten vier Töne zu Beginn des Saxophon-Solos; diese werden teils durch den ausklingenden Gesang, teils durch das E-piano im Frequenzbereich bis 500 Hz verdeckt. Die erste Harmonische des Gesangs (Vokal /a/, Grundfrequenz 600 Hz) ist nur noch schwach ausgeprägt erkennbar.

Die Resynthese der Saxophonstimme gibt dennoch einen guten Eindruck vom Melodieverlauf, was auf das Entstehen von virtuellen Tonhöhen hinweist. Die Anwendung des Tonhöhenberechnungsverfahren nach [7] auf die extrahierte Saxophonstimme liefert die in Abb.3 rechts dünn gezeichneten Tonhöhen. Außer für den ersten Ton, dessen Spektralkomponenten nicht zu einer signifikanten virtuellen Tonhöhe führen, wird die in der extrahierten Saxophonstimme fehlende Grundfrequenz ermittelt.

Zieht man das TTZM der extrahierten Saxophonstimme von Gesamt-TTZM ab, erhält man ein von der Saxophonstimme befreites Signal. Die Resynthese dieses Rest-TTZM enthält nur noch Gesang und E-piano, das Saxophon ist nicht mehr wahrnehmbar. Der Klangeindruck, den das Rest-Signal hinterläßt ist aber gegenüber dem Gesamtsignal stark gestört. Das bedeutet, daß Spektralanteile der Saxophonstimme zum Klangeindruck des Gesangs beigetragen haben müssen. Hier zeigen sich die Grenzen des Extraktionsverfahrens: Die Frage, in welchem Maße Spektralkomponenten eines „Maskierers“ (Saxophon) zum Klangfarbeneindruck eines teilweise maskierten Schalles (Gesang) beiträgt, bedarf weiterer Untersuchungen.

Literatur

- [1] Askenfelt, A., Automatic notation of played music: the Visa project. *Fontes Artis Musicae* 1979,16(2),109-120.
- [2] Baumann, U., Stimmentrennung in polyphoner Musik. In *Fortschritte der Akustik (DAGA 92)*. Bad Honnef - Bochum, 1992, S. 725 -728.
- [3] Heinbach, W., Aurally adequate signal representation: The Part-Tone-Time-Pattern. *Acustica* 1988, 67, 113-121.
- [4] Katayose, H. et al., Sentiment Extraction in music. *Proc. of ICASSP* 1988,1083-1087.
- [5] Loosen, F., Intonation of solo violin performance with reference to equally tempered,Pythagorean, and just intonations. *J.Acoust.Soc.Am.* 1993,93(1),525-539.
- [6] Repp, B., Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei". *J.Acoust.Soc.Am.* 1992,92(5),2546-2568.
- [7] Terhardt, E., Calculating virtual pitch. *Hearing Research* 1979, 1,155-182.