

Anwendungsbeispiele für Musikanalyse-Algorithmen

Janto Skowronek, Martin McKinney, Steven van de Par, Jeroen Breebaart

Philips Research Laboratories Eindhoven, Email: janto.skowronek@philips.com

Einleitung

Seit einigen Jahren werden vielfältige Algorithmen zur automatischen Analyse von Musik (hier "Musiktechnologie" genannt) entwickelt. Dieser Beitrag gibt einen kurzen Abriss des potentiellen Marktes für Musiktechnologie sowie eine Übersicht der Funktionen, die die Algorithmen dort übernehmen können. Anschliessend werden vier eigene Ansätze als Beispiialgorithmen präsentiert.

Markt- und Technologieübersicht

Die wichtigsten Anwendungsfelder für Musiktechnologie sind die Internet basierte Musikdistribution und die Endbenutzerprodukte. Bezüglich der Musikdistribution über das Internet gibt es drei Dienste: Download, Streaming und Sharing. Die größten Download-Anbieter sind iTunes, Rhapsody, Napster, Amazon und Yahoo, die größtenteils auch Streaming und Sharing anbieten. Streaming-Dienste wie LastFM, Pandora, SocialFM und Musicoverly konkurrieren mit den "gewöhnlichen" Internet-Radiostationen durch Anbieten von personalisierten Internetradios. Letztendlich gibt es auch legale Sharing-Dienste wie z.B. Napster, Grooveshark und QTrax. Die Endbenutzerprodukte kann man grob in PC-Anwendungen, Mobiltelefon-Anwendungen, tragbare Musikspieler und Standalone-Geräte einteilen. Neben den Standard-Musikplayern für den PC (z.B. MediaPlayer, Winamp) und Frontend-Software der oben genannten Dienste (z.B. iTunes) gibt es auch weitere Player (z.B. Moodlogic) die mittels Musiktechnologie erlauben, effizient die eigene Musikkollektion zu organisieren und zu hören. Bei den Mobiltelefonen können die Geräte aller grossen Hersteller Musik abspielen. Erwähnenswert wäre hier Sony-Ericsson, die es erlauben, Musik abhängig von der Stimmung zu suchen. Download- oder Streaming-Dienste speziell für Mobiltelefone bieten z.B. Shazam, SocialFM und Beatbrew an. Bei den tragbaren Musikspielern übernehmen oft PC-basierte Anwendungen (z.B. iTunes, Moodlogic) die komfortable Musikorganisation. Die Spieler selbst bieten meist nur einfache Skip- und Zufallswiedergabe-Funktionen. Bei den Standalone-Geräten wie Heimkinos, HiFi-Systeme und MediaCenter heben sich einige Hersteller mit fortgeschrittenen Musikmanagement aus dem Markt hervor. Die Philips Wireless-Music Geräte z.B. verwenden Gracenotes Musik-Identifizierungs-Technologie zur Musikorganisation mittels Genres.

Der Einsatz von Musiktechnologie kann 5 Funktionen unterstützen: Suche (Wiederfinden, Neues entdecken), Zusammenstellung (für einen Zeitraum), Teilen (Musikgeschmack, Musiktitel), Manipulation (Remix, AutoDJ, etc.) und Verbindung mit Zusatzmaterial (z.B.

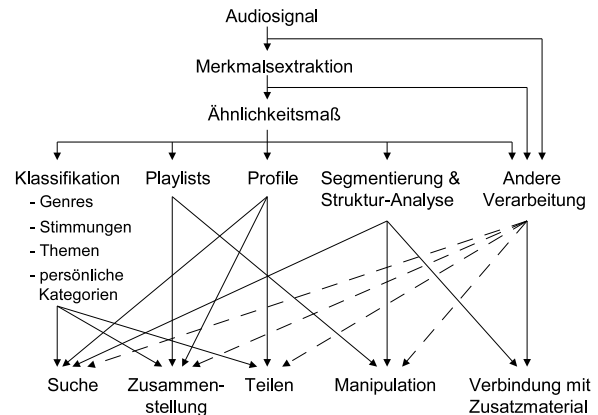


Abbildung 1: Übersicht Musikanalyse Technologie.

Texte, Notation). Abbildung 1 illustriert die Hauptverarbeitungsschritte vom Musiksignal zur Funktionalität: Gewöhnlich werden von den Audiosignalen charakteristische Merkmale extrahiert, mit denen ein Abstandsmaß zwischen Musiktiteln berechnet werden kann. Mittels eines solchen Abstandsmaßes kann dann Musik klassifiziert werden, Playlists oder Benutzerprofile erstellt werden, oder eine Segmentierung und Strukturanalyse einzelner Musiktitel durchgeführt werden. Ferner können andere Verarbeitungen sowohl Audiosignal, Merkmale oder Abstandsmaß verwenden. Musikklassifikation und Profile können bei der Suche, der Zusammenstellung und dem Teilen von Musik helfen. Playlistgenerierung ermöglicht die Musikzusammenstellung und -manipulation (z.B. automatischer Mix). Segmentierung und Strukturanalyse unterstützt bei der Suche (z.B. Auffinden des Refrains), der Manipulation (z.B. gute Mix-Punkte), und der Verbindung mit Zusatzmaterial (z.B. Karaoke). Andere Verarbeitungen von Musik sind für alle Funktionen denkbar, vornehmlich aber eher für die Verbindung mit Zusatzmaterial (z.B. Synchronisierung von Audiosignal und Notation).

Beispiele für Musikanalyse-Algorithmen

1) Musikgenre Klassifikation

Es gibt sehr viele Arbeiten zur automatischen Klassifikation von Musikgenres, eine Übersicht gibt z.B. [1]. In dem eigenen Ansatz zur Genre-Klassifikation lag der Schwerpunkt auf der Entwicklung eines geeigneten Merkmalsatzes und der Zusammenstellung einer angemessenen Test- und Trainingsdatenbasis. Der verwendete Merkmalsatz kombiniert spektro-temporale Merkmale [2] mit musikalisch basierten Merkmalen, die die tonale Struktur der Musik [3], das Auftreten perkussiver Klänge in der

Musik [4] sowie Rhythmus und wahrgenommenes Tempo [5] beschreiben. Die ca. 1700 Titel umfassende Datenbasis für die Genreklassifikation definiert 12 Musikstile westlicher Musik. Die Klassen sowie die dazugehörigen Titel wurden von Experten zusammengetragen unter der Vorgabe, nur prototypische, d.h. keine Misch-Stile, und disjunkte, d.h. sich gegenseitig ausschliessende, Genre-Klassen zu bilden. Als Klassifikationsmethode kam eine Standardmethode (Quadratische Diskriminanz-Analyse) zum Einsatz und mittels einer 80/20 Training/Test-Aufteilung und randomisierten Wiederholungen wurde die Klassifikationsleistung ermittelt, die für die 12 Genres bei ca. 71 % korrekter Erkennung lag.

2) Klassifikation der Stimmung in Musik

Bei der Entwicklung eines Klassifikators für Musik-Stimmungen spielt Subjektivität eine entscheidende Rolle. Daher wurde in einem experimentellen Ansatz [6] diese Subjektivität minimiert, um die größtmögliche Akzeptanz über Hörer hinweg zu erreichen: a) Es wird modelliert, welche Stimmung die Musik ausdrückt und nicht wie sich der Hörer tatsächlich fühlt. b) Nur die Klassen werden modelliert, die durch Versuchspersonen relativ übereinstimmend verwendet werden. c) Die Trainings- und Testdatenbasis beinhaltet nur Musik, welche Versuchspersonen deutlich den Stimmungsklassen zuordnen können. Ergebnis dieses Verfahrens ist eine Datenbasis von ca. 1000 Musiksegmenten, für die eine Beurteilung durch Versuchspersonen in 12 Stimmungsklassen vorliegt. Da die Klassendefinition eine Überschneidung von Klassen ermöglichte, wurden für jede Stimmung individuelle Detektoren unter Verwendung der o.g. Klassifikationsmethode und o.g. Merkmale trainiert. Je nach Stimmungskategorie wurde 77 - 91 % korrekte Erkennung erreicht. Eine subjektive Qualitätsbeurteilung für 180 neue, in der Datenbasis nicht verwendete Musiksegmente ergab, dass in 80 % der Fälle die algorithmische Schätzung "angemessen" oder besser war, in 60 % "gut" oder besser.

3) Musik-Ähnlichkeitsberechnung

Die Berechnung eines Abstandsmaßes zwischen zwei Merkmalsvektoren - und damit der Musik, die sie repräsentieren - ist eine Kernoperation von Musikanalysealgorithmen. Eine Möglichkeit eines solchen Abstandsmaßes ist die sog. Mahalanobis-Distanz $D = (f - g)^T W (f - g)$ mit f, g = Merkmalsvektoren und W = inverse Kovarianzmatrix aller vorhandenen Merkmalsvektoren. Dieses Abstandsmaß beinhaltet mit W eine Gewichtung, um die unterschiedliche Skalierung von Merkmalen zu berücksichtigen. Um jedoch überzeugende Ergebnisse zu liefern, muss ein solches Abstandsmaß den "wahrgenommenen Abstand" zweier Musiktitel modellieren. Während dazu eine eigene perzeptive Studie läuft [7], wurde für den gegenwärtigen Algorithmus eine Gewichtsmatrix mittels einer Datenbasis so ermittelt, dass das Abstandsmaß D zwischen zwei Musikstücken klein ist wenn beide zu dem selben Musik-Genre gehören oder wenn beide von demselben Künstler stammen. Mittels eines solchen Abstandsmaßes lassen sich dann anschließend einfache Playlists erstellen: Der Benutzer wählt einen "Start-Song" und der

Algorithmus arrangiert die z.B. 20 ähnlichsten Musiktitel aus der Musikkollektion in einer Playlist.

4) Klassifikation persönlicher Kategorien

Als letztes Beispiel soll die automatische Einteilung von Musik in von dem Benutzer definierte "persönliche Kategorien" vorgestellt werden. Die Herausforderung eines solchen Algorithmus ist das schnelle Erreichen zufriedenstellender Resultate. Der Algorithmus muss sofort angemessene Ergebnisse liefern, ohne dass der Benutzer den Klassifikator in einer langwierigen Prozedur selbst trainieren muss. Eine gängige Lösung ist es, von einer vorab definierten Basiskonfiguration ausgehend in einem iterativen Prozess die Klassifikationsleistung zu personalisieren. Dabei wird nach erfolgter Klassifikation dem Benutzer die Möglichkeit gegeben, fehlerhaft eingeteilte Musiktitel zu korrigieren. Anschliessend wird das Klassifikationsmodell aktualisiert und die Klassifikation erneut durchgeführt. Dazu verwendet der eigene Ansatz ein vorab definiertes "allgemeines Modell", welches die einzelnen Merkmale angemessen gewichtet. Diesem wird dann ein "spezielles Modell" aufgesetzt, das die Klassenwahrscheinlichkeitsfunktionen aufgrund von Beispielsongs berechnet, welche der Benutzer angibt um seine persönlichen Klassen zu definieren. Wenn der Benutzer Korrekturen der Klassifikationsresultate vornimmt, werden die Wahrscheinlichkeitsfunktionen in dem "speziellen Modell" optimiert.

Literatur

- [1] P. Ahrendt, *Music Genre Classification Systems*, PhD dissertation, Technical University of Denmark, Kongens Lyngby, Denmark, 2006.
- [2] M. McKinney, J. Breebart, *Features for Audio Music Classification*, 4th International Conference on Music Information Retrieval, 2003.
- [3] S. van de Par, M. McKinney, A. Redert, *Musical Key Extraction from Audio using Profile Training*, 7th International Conference on Music Information Retrieval, Victoria, 2006.
- [4] J. Skowronek, M. McKinney, *Perkussivität von Klängen als neues Merkmal zur Klassifikation von Musik-Genres*, 32. Jahrestagung für Akustik, Braunschweig, 2006.
- [5] M. McKinney, D. Moelants, *Extracting the perceptual tempo from music audio*, 5th International Conference on Music Information Retrieval, Barcelona, 2004.
- [6] J. Skowronek, S. van de Par, M. McKinney, *Erstellen einer Trainings- und Test-Datenbasis für die Klassifikation der Stimmung in Musik*, 33. Jahrestagung für Akustik, Stuttgart, 2007.
- [7] A. Novello, M. McKinney, *Assessment of perceptual music similarity*, 8th International Conference on Music Information Retrieval, Vienna, 2007.

Introduction

Since several years various algorithms for the automatic music analysis (here called “music technology”) are developed. This conference contribution will give a short overview on the potential market for music technology as well as the functionalities that the algorithms could enable. Then a presentation of four own approaches as examples for such algorithms will conclude the paper.

Market and Technology Overview

The main application fields for music technology are the Internet based music distribution and the end consumer products. Internet music distribution comprises three forms of services: Download, Streaming & Sharing. The largest download services are iTunes, Rhapsody, Napster, Amazon & Yahoo, which for the most part provide streaming and sharing as well. Streaming services such as LastFM, Pandora, SocialFm and Musicoverly compete the the “normal” Internet radios by offering personalized radio streams. Eventually there are also legal sharing services such as Napster, Grooveshark and QTrax. The end consumer products can be roughly divided into PC applications, mobilephone applications, portable music player and stand-alone devices. Next to the standard music players for PC (e.g. MediaPlayer, Winamp) and frontend software of above mentioned services (e.g. iTunes) are also other players (e.g. Moodlogic) that allow an efficient organisation and listening to the own music collections by means of music technology. With respect to the mobilephones, devices of all big producers can replay music. Here Sony-Ericsson is worth mentioning, because they allow the search for music by the music’s mood. Download or streaming services dedicated to mobilephone are for instance offered by Shazam, SocialFM and Beatbrew. With respect to the portable devices, often PC applications (e.g. iTunes, Moodlogic) take over the comfortable music organisation, while the partables themselves often support only skip and random replay functionality. Among the stand-alone devices such as HomeCentres, HiFi systems and MediaCenters, several manufacturers divide from the market with advanced music management functionalities. The Philips Wireless-Music devices for instance use Gracenote’s Music-Identification-Technology to enable a music organization using genres.

Using music technology can support 5 main functionalities: search (find back, discover new), compiling (for a time period), sharing (music taste, music tracks), manipulation (remix, AutoDJ etc.) and Connection with editorial material (e.g. lyrics, score notation). Figure 1 depicts the main processing steps from music signal to functionality: Usually characteristic features are extracted from the audio that enable then a computation of a distance measure between music tracks. Then such a distance measure enables the classification of music, the generation of playlists or user profiles, or a segmentation and structure analysis of music tracks. Furthermore other processing steps are possible that could use the au-

dio signal, the features or the distance measure. Music classification and profiles can help with the search, teh compilation and the sharing of music. Playlist generation enables the compilation of music and also manipulation (e.g. automatic mix). Segmentation and structure analysis support search (e.g. by finding chorus), manipulation (e.g. good mixing points) and the connection with editorial content (e.g. karaoke). Other music processing steps are imaginable for all functions, but they are in particular interesting for the connection with editorial content (e.g. audio score synchronization).

Examples for Music Analysis Algorithms

1) Music Genre Classification

Much work is reported on teh automatic classification of music genres, an overview is given in [1]. The own approach for genre classification focussed on the development of a suitable feature set and the compilation of a proper ground-truth database. The used feature set combines spectro-temporal features [2] with musicologically based features that describe the tonal structure of the music ([3], the occurrence of percussive events in the music [4] as well as rhythm and perceived tempo [5]. The compiled genre ground-truth database defines 21 music genres of western music and comprises ca. 1700 tracks. Experts selected the classes and corresponding tracks given the requirements that only prototypical - no fusion genres - and mutually exclusive genre classes should be compiled. We applied a standard classification technique (quadratic discriminant analysis) and using a 80/20 training/test split with randomized repetitions we estimated the classification performance. For the 12 genres correct classification rate of 71 % could be achieved.

2) Music Mood Classification

When developing a music mood classifier, one has to treat the subjectivity involved in this task. Therefore we minimized that subjectivity in an experimental approach [6] to achieve a maximum acceptance across people: a) The algorithm models the mood the music conveys and not how the listener would actually feel. b) Only those mood classes are modelled on which subjects showed a relative agreement when using them. c) The ground-truth database should only contain music that subjects can clearly assign to the mood classes. Result of this procedure is a database of ca. 1000 music excerpts, for which a assessment by subjects are available for 12 mood classes. Since the mood classes were not defined as mutually exclusive, for every mood an individual detector was trained using above mentioned classification method and feature set. Depending on the mood class, the achieved correct recognition rate was between 77 - 91 %. a subjective quality evaluation for 180 excerpts that were not part of the ground-truth database showed that in 80% of the cases the algorithmic mood predictions were “fair” or better, in 60 % “good” or better.

3) Music Similarity Measurement

The computation of a distance measure between two feature vectors - and with that the music represented by them - is a core operation of music analysis algorithms. One possibility for such a distance measure is the Mahalanobis distance $D = (f - g)^T W (f - g)$ with f, g = feature vectors and W = inverse covariance matrix of all available feature vectors. This distance measure constrains with W a weighting that accounts for different scaling of features. But in order to deliver convincing results, such a distance measure should model the “perceived distance” between two music tracks. While an own perceptual study on perceptual music similarity is currently performed [7], the current algorithm uses a weighting matrix that has been computed with a music database such that the distance measure between two music tracks is small if both pieces are from the same genre or from the same artist. Finally such a distance enables to generate simple playlists: The user chooses a seed song and the algorithm compiles a playlist with the for instance 20 most similar tracks in the user’s music collection.

4) Personal Categories Classification

The last example that is presented here is the automatic classification of music into user defined “personal categories”. The challenge of such an algorithm is a quick achievement of satisfactory results without a large effort required from the user. That means the algorithm has to deliver immediately adequate results without demanding the user to train the classifier in a tedious procedure. A common solution is to personalize the classifier starting from a basic configuration. In that process the user gets the possibility to correct false class assignments of tracks and the system will update the classification model and reclassify the music. The own approach uses a pre-defined “generic model” that accounts for a proper weighting of features. On top of that, a “specific model” computes the class probability functions using example songs that the user has to give for defining his personal categories. If the user is correcting the classification results, the class probability functions in the “specific model” will be optimized.