

Fachhochschule Potsdam  
Fachbereich Informationswissenschaften

## **Die Bedeutung und Anwendung von Music Information Retrieval**

Diplomarbeit zur Erlangung des akademischen Grades:  
Diplom-Dokumentar(in) FH

eingereicht von

Christine Höfer

Matrikelnummer:	6757
Studiengang:	Dokumentation
Anschrift:	Christine Höfer Hinter der Mauer 8 07318 Saalfeld
Gutachter:	Prof. Dr. habil. Ralf-Dirk Hennings Fachbereich Informationswissenschaften  Dipl.-Ing. (FH) Christian Dittmar Fraunhofer-Institut für Digitale Medientechnologie
Abgabetermin:	30. Juni 2009

---

## **Eidesstattliche Erklärung**

„Ich versichere hiermit an Eides statt, dass ich die vorliegende Diplomarbeit selbständig verfasst und hierzu keine anderen als die angegebenen Hilfsmittel verwendet habe. Alle Stellen der Arbeit, die wörtlich oder sinngemäß aus fremden Quellen entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form in keinem anderen Studiengang als Prüfungsleistung vorgelegt oder an anderer Stelle veröffentlicht. Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.“

Saalfeld, 29. Juni 2009

---

Christine Höfer

---

# Inhaltsverzeichnis

Einleitung .....	8
Kapitel 1 : Definition von Music Information Retrieval.....	10
1. Definition und Zweck von Music Information Retrieval.....	10
1.1 Die 7 Facetten von Musikinformation nach J. Stephen Downie.....	11
Kapitel 2 : Musiklehre .....	13
2. Grundlegende Begriffe .....	13
2.1 Basisdimensionen .....	13
2.2 Zusätzliche Begriffe der Musiklehre .....	17
Kapitel 3 : Repräsentation von Musik .....	22
3. Musikrepräsentation .....	22
3.1 Audio-Repräsentation.....	23
3.2 Symbolische Repräsentation .....	29
Kapitel 4 : Audiostandards .....	33
4. Standards für den Audio Bereich.....	33
4. 1 Moving Pictures Expert Group .....	33
Kapitel 5 : Merkmalsextraktion .....	39
5. Grundlagen der Merkmalsextraktion.....	39
5.1 Fourier Transformation .....	39
5.2 Wavelet Transformation.....	41
5.3 Diskrete Kosinus Transformation.....	42
5.4 Hauptkomponentenanalyse .....	43
5.5 Merkmale der Merkmalsextraktion .....	44
Kapitel 6 : Mustererkennung.....	50
6. Verfahren zur automatischen Mustererkennung .....	50
6.1 Überwachte Lernverfahren .....	51
6.2 Transformationen im Merkmalsraum .....	54
Kapitel 7 : Musiksynchronisation .....	56
7.1 Bedeutung von Musiksynchronisation für Musikbibliotheken .....	56
7.2 Audio-Partitur Synchronisation (audio-to-score alignment) .....	57
7.3 Dynamic Time warping .....	58
7.4 SyncPlayer System .....	59
Kapitel 8 : Suchmethoden .....	60
8.1 Überblick der Hauptsuchmethoden des Audio Retrieval .....	60
8.2 Einzelne Suchmethoden.....	62
Kapitel 9: Anwendungen .....	70
Anwendungen von Music Information Retrieval im Internet .....	70
9.1 Musiksuchmaschinen in der Praxis.....	70
9.2 Technologien von Musikempfehlungssystemen.....	79
9.3 Beispiele für Empfehlungssysteme.....	81
9.4 Übersicht von einigen Portalen mit grober Funktionskategorisierung.....	86
Schluss / Zusammenfassung .....	88

---

Anhang .....	89
Quellenverzeichnis.....	98

---

## Abkürzungsverzeichnis

CB	Content-Based
CF	Collaborative Filtering
CMN	Common Music Notations
DCT	Discrete Cosine Transformation
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transformation
FT	Fourier Transformation
GMM	Gaussian Mixture Model
kNN	k-Nearest-Neighbour
LDA	Linear Discriminant Analysis
MFCC	Mel Frequency Cepstral Coefficients
MIDI	Musical Instrument Digital Interface
MIR	Music Information Retrieval
NIFF	Notation Interchange File
PC	Parsons Code
PCA	Principal Component Analysis
PCM	Pulse-Code-Modulation
QbE	Query by Example
QbH	Query by Humming
QbM	Query by Melody
QbT	Query by Tapping
RMS	Root Mean Square
SC	Spectral Centroid
SR	Spectral Rolloff
SVM	Support Vector Machine
ZCR	Zero Crossing Rate

---

## Abbildungsverzeichnis

Abbildung 1: wissenschaftliche Felder von MIR.....	10
Abbildung 2 : Zusammenhang zwischen Zahlenwert der Frequenzangabe und der empfundenen Tonhöhe .....	14
Abbildung 3 : Akkord .....	15
Abbildung 4 : Einfache Intervalle der Musik.....	19
Abbildung 5 : Notensystem.....	21
Abbildung 6 : Eigenschaften einer einzelnen Note .....	21
Abbildung 7 : grobe Übersicht zur Musikrepräsentation.....	22
Abbildung 8 : Darstellung von Musikformaten .....	22
Abbildung 9 : harmonische Sinusschwingung .....	23
Abbildung 10 : Abtastung eines Audiosignals.....	25
Abbildung 11 : Darstellung der Quantisierung.....	25
Abbildung 12 : Darstellung einer Wellenform .....	26
Abbildung 13 : piano-roll .....	29
Abbildung 14 : Ausschnitt eines MusicXML Samples.....	31
Abbildung 15 : tragende Rolle des MPEG-7 Standard in der Musikverarbeitung des Internet .....	37
Abbildung 16 : Signal vor der Transformation .....	42
Abbildung 17 : nach der Transformation / Spitzen zeigen Cosinus Frequenzen .....	43
Abbildung 18 : Beispiel Beat-Histogram (BPM=beat per minute).....	48
Abbildung 19 : Beispiel Chromagramm – Frequenzverlauf (x-Frequenz, y- Oktaven) .....	49
Abbildung 20 : Schritte von einem Audio-Signal zu einem Chromagramm...	96
Abbildung 21 : Support Vector Machines (einfache Darstellung) .....	53
Abbildung 22 : Verlinkung von Musikdaten in verschiedenen Formaten .....	57
Abbildung 23 : Auszug der Genre von Musuma.....	65
Abbildung 24 : Prozesse einer Similarity Search .....	68
Abbildung 25 : Klaviatur Suche von Musipedia .....	71
Abbildung 26 : Eingabe des Parsons Code .....	72
Abbildung 27 : Metadatenuche von Musicline .....	74
Abbildung 28 : Genresuche mit musicline.....	75
Abbildung 29 : Eingabe der Melodiesuche .....	76
Abbildung 30 : Ergebnis Test 1 - Musicline.....	76
Abbildung 31 : Ergebnis Test 2 - Musicline.....	76
Abbildung 32 : Melodiesuche von Midomi .....	77
Abbildung 33 : Ergebnis Test 1 - Midomi .....	78
Abbildung 34 : Ergebnis Test 2 - Midomi .....	78
Abbildung 35 : Ausschnitt des Empfehlungsbereiches von amazon.....	82
Abbildung 36 : „Ähnliche Künstler“ Funktion bei Last.fm.....	84
Abbildung 37 : Visualisierung des Musikgeschmacks von musicovery .....	85

---

## Tabellenverzeichnis

Tabelle 1 : Liste von diversen Audioformaten.....	28
Tabelle 2: Symbole des Parsons Code.....	32
Tabelle 3 : Überblick von MPEG-Standards.....	34
Tabelle 4 : Überblick von Merkmalen (nicht alle Merkmale werden in dieser Arbeit beschrieben).....	44
Tabelle 5: überwachte und unüberwachte Lernverfahren .....	50
Tabelle 6 : Beispiele für den Parsons Code.....	64
Tabelle 7 : einfache Übersicht von Musikempfehlungssystemen und Musiksuchmaschinen .....	87

## Einleitung

Die vorliegende Arbeit beschäftigt sich mit der Thematik „*Music Information Retrieval*“. Der Begriff „*Information Retrieval*“ wird häufig mit dem herkömmlichen Textretrieval verbunden, obwohl zum jetzigen Zeitpunkt Recherchen auch ohne Text möglich sind. Viele Jahre der Entwicklung haben dazu geführt, dass Informationen anhand von Ton und Zeichnungen recherchiert werden können. Für die Erschließung von Dokumenten wird generell auf Metadaten zur eindeutigen Identifikation zurückgegriffen. Das ist soweit auch für Audiodateien möglich, aber im Gegensatz zu normalen Textdokumenten, besitzen sie wesentlich mehr charakteristische Merkmale. Jede Melodie hat beispielsweise eine eigene Tonhöhe oder einen eigenen Rhythmus und viele weitere Faktoren, die den Song einzigartig machen. Die Schwierigkeit besteht in dem Extrahieren von solchen Informationen, die später für Recherchen verwendet werden. Der Schwerpunkt der vorliegenden Arbeit umkreist die vorhandenen Möglichkeiten der Nutzung von Music Information Retrieval im Internet. Es existiert eine Vielfalt an Musikempfehlungssystemen mit denen Internetnutzer häufig konfrontiert werden. Beispielsweise bei dem Kauf einer Audio - CD des Online-Shops „*amazon*“ wird anhand des Kaufverhaltens des Kunden eine Empfehlung für weitere Artikel erstellt. Beinahe jede dieser Musikempfehlungen basiert auf einem bestimmten Algorithmus von MIR. Die Methoden sind aber noch wesentlich vielfältiger. Man könnte sich folgende Problematik vorstellen. Eine so genannte „*Ohrwurm-Melodie*“ kann wochenlanges Kopfzerbrechen bereiten, wenn der Titel unbekannt ist. Um dieses Problem zu lösen, wurde das „*Query by Humming*“ - System entwickelt. Der Nutzer ist damit beispielsweise in der Lage, die bekannte Melodie als Anfrage durch Summen, Pfeifen oder Singen in ein Mikrofon zustellen. Dabei spielt es keine Rolle, ob es sich um den Anfang, die Mitte oder das Ende des Stückes handelt. Das System vergleicht den Bruchteil des gesuchten Liedes anhand verschiedener Komponenten, mit den bereits vorhandenen Stücken in der Datenbank und listet die Ergebnisse auf. Weiterhin ist auch eine Anfrage durch das „*Tappen*“ des Rhythmus möglich. Doch zunächst beschäftigt sich die Arbeit mit den Grundlagen von MIR, um einen kompakten Überblick zu



---

verschaffen, wie diese Methoden entstehen und welche mathematischen Verfahren zum Tragen kommen. Zum Verständnis dieser komplexen Vorgänge ist ein gewisses Grundlagenwissen aus dem Bereich Musiktheorie bzw. Musiklehre unerlässlich.

Dabei stehen zunächst die Basiselemente von Musik, wie beispielsweise Tonhöhe und Klangfarbe, im Mittelpunkt. Diese spielen eine primäre Rolle bei den Aspekten des Musik-Retrieval. Das Kapitel „Merkmalsextraktion“ beschreibt die mathematischen Verfahren der Signalverarbeitung die für MIR von Bedeutung sind. Dabei ist es zusätzlich nützlich zu wissen, wie diese Signale klassifiziert werden. Die dazugehörigen Klassifikationsverfahren werden im Kapitel „Mustererkennung“ näher vorgestellt. Somit ist dann eine Grundlage geschaffen, für den Übergang zu den Hauptsuchmethoden und Anwendungen.

Das Hauptziel der Arbeit ist es, einen überschaubaren Einblick in diesen hochkomplexen Bereich zu gewähren und die Möglichkeiten dieser fortschrittlichen Technik darzustellen. Neue Aspekte und Technologien werden auf zahlreichen Veranstaltungen wie dem *International Symposium on Music Information Retrieval*; *DafX Conference*; *ICASSP - International Conference on Acoustics, Speech, and Signal Processing*; *ICMC – International Computer Music Conference etc.* vorgestellt. Somit nehmen die Verfahren stetig an Effektivität und Präzision zu und vereinfachen dem Nutzer die Suche.

## Kapitel 1 : Definition von Music Information Retrieval

### 1. Definition und Zweck von Music Information Retrieval

Music Information Retrieval ist eine vergleichsweise junge Disziplin, obwohl sie bereits seit den 1960er<sup>1</sup> Jahren existiert und mit vielen verschiedenen wissenschaftlichen Bereichen verknüpft ist. (siehe Abb.1).<sup>2</sup>

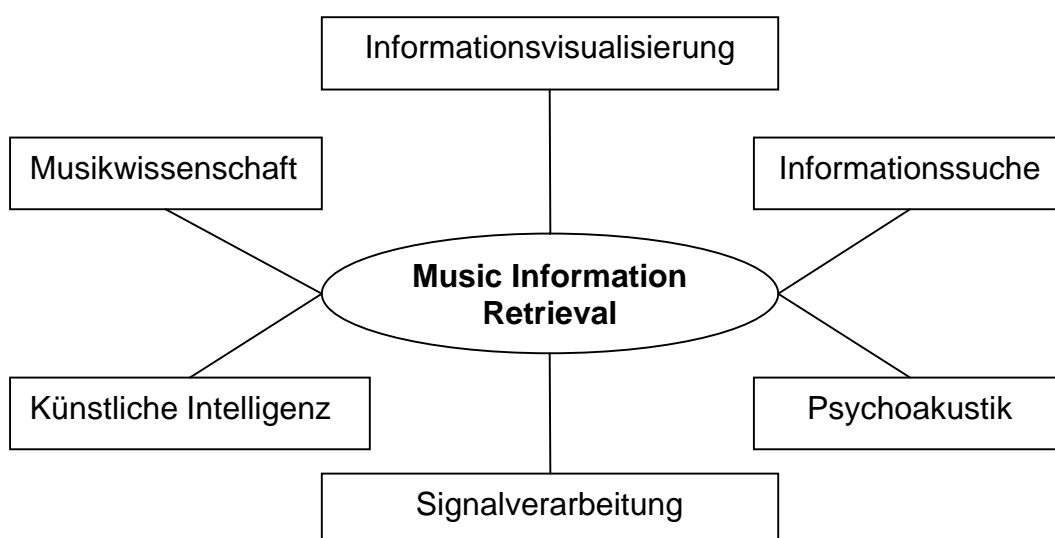


Abbildung 1: wissenschaftliche Felder von MIR

Die Notwendigkeit von *Music Information Retrieval*, erschließt sich durch einen Blick auf die umfangreichen Musikkollektionen im Internet. Früher nur mit Metadaten erschlossen, haben sich auf Basis von MIR neue vielfältige Wege entwickelt. Das Ziel besteht darin, Technologien bzw. Werkzeuge zu konstruieren, welche einen effizienten Zugriff auf Musikinformationen gewähren. Die Kernaufgaben bestehen in der Analyse, Repräsentation, Modellierung, Klassifizierung und Clustering von Musik<sup>3</sup>. Es ist notwendig semantischen Beziehungen zwischen den Stücken zu analysieren, die sich

<sup>1</sup> Vgl. Music Information Retrieval. <http://www.dfki.de/web/forschung/km/kompetenz/forschung/music-information-retrieval> [letzter Zugriff: 02.05.2009]

<sup>2</sup> Vgl. Music Information Retrieval. <http://futurezone.orf.at/stories/275698/> [letzter Zugriff: 02.05.2009]

<sup>3</sup> Vgl. Automatische Klassifikation von Musikkünstlern basierend auf Web-Daten (2005). [http://www.cp.jku.at/research/papers/knees\\_thesis.pdf](http://www.cp.jku.at/research/papers/knees_thesis.pdf) [letzter Zugriff: 10.06.2009]

beispielsweise rhythmisch oder harmonisch äußern können. Hierbei spielt auch die Komplexität von Musik eine tragende Rolle. Ein klassisches Stück, wie eine Symphonie von Beethoven, weist wesentlich mehr Komplexität auf, als ein herkömmlicher Popsong und kann demnach erheblich schwerer analysiert werden. Der Vorteil besteht aber darin, dass solche Musik mehr eindeutige Merkmale aufweist. Ein Mensch könnte diese beiden Musikgattungen einfach unterscheiden, hingegen eine Maschine eine solche akustische Sensibilität nicht aufweisen kann. Das System orientiert sich darum an der Struktur von Musik, den Bestandteilen. Diese geben Aufschluss darüber, wie ein Song sich in seinen Eigenschaften von einem anderen unterscheidet.

Music Information Retrieval umfasst folgende Teilbereiche<sup>4</sup>:

1. „Verarbeitung von Musik, die in symbolischer Form vorliegt
2. Verarbeitung von Musik, die als Signal vorliegt
3. Verarbeitung von kulturellen Informationen (community metadata)
4. Benutzer-orientierte Entwicklung von Schnittstellen“<sup>5</sup>

### 1.1 Die 7 Facetten von Musikinformation nach J. Stephen Downie<sup>6</sup>

Downie stellt Musikinformation in 7 Facetten dar<sup>7 8</sup>.

- Pitch (Tonhöhe)
- Temporal (Tempo)
- Harmonic (Harmonik)
- Timbral (klangfarblich)
- Editorial (redaktionell)
- Textual (textlich)
- Bibliographic (bibliographisch)

---

<sup>4</sup> Automatische Klassifikation von Musikkünstlern basierend auf Web-Daten (2005).  
[http://www.cp.jku.at/research/papers/knees\\_thesis.pdf](http://www.cp.jku.at/research/papers/knees_thesis.pdf) [letzter Zugriff: 10.06.2009]

<sup>5</sup> keine scharfe Grenzen, Bereiche können auch kombiniert werden

<sup>6</sup> Dozent, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign,

<sup>7</sup> Vgl. Annual Review of Information Science and Technology 37, 295-340

<sup>8</sup> Facette 1-4 wird in Kapitel 2 näher erläutert

### 1.1.1 Redaktionelle Facette

Die redaktionelle Facette setzt den Fokus auf die Beschreibung der Verfeinerung eines musikalischen Stückes. Hierzu zählen die Fingersätze, die Verzierungen, dynamische Instruktionen, Artikulation, Bindebögen, Bogenführung usw.. Eine Schwierigkeit bei der redaktionellen Facette ist, dass sie symbolisch (-,3,!) oder textlich (z.B. crescendo, diminuendo) sein kann, sowie auch beides. Weiterhin kann die redaktionelle Facette auch Fragmente der Musik selbst beinhalten, beispielsweise anhand des Herausschreibens von Harmonien. Die redaktionellen Diskrepanzen zwischen den verschiedenen Versionen desselben Werkes, erschweren die Wahl der „definitiven“ Version für die Aufnahme in ein MIR System.<sup>9</sup>

### 1.1.2 Textliche Facette

Die textliche Facette bezieht die Lyrik von Liedern, Hymnen, etc. mit ein. Sie ist weitestgehend unabhängig von Melodien und musikalischen Arrangements. Manchmal ist ein lyrisches Fragment nicht informativ genug, um die gewünschte Melodie zu identifizieren. Ein Lied kann mehrere Texte und ein Text mehrere musikalische Varianten aufweisen. Des Weiteren ist zu beachten, dass Musik nicht unbedingt Lyrik aufweisen muss.<sup>10</sup>

### 1.1.3 Bibliographische Facette

Die bibliographische Facette widerspiegelt Informationen über den Titel, Komponist, Songschreiber, Publisher usw.. Es handelt sich ausschließlich um Metadaten eines Musikstückes. Die bibliographische Facette ist die einzige Facette, die sich nicht vom Inhalt einer Komposition ableiten lässt, es sind Informationen über ein musikalisches Werk, die deskriptiv erfasst werden.<sup>11</sup>

---

<sup>9</sup> Vgl. Annual Review of Information Science and Technology 37, 295-340

<sup>10</sup> ebenda

<sup>11</sup> ebenda

## Kapitel 2 : Musiklehre

### 2. Grundlegende Begriffe

Um die komplexen Beziehungen von MIR verstehen zu können, ist ein grundlegendes Wissen aus der Musiklehre unerlässlich. Dieses setzt sich aus verschiedenen Lehren (Notenlehre, Harmonielehre usw.) über Musik zusammen.

Das Musikretrieval befasst sich vorwiegend mit folgenden 4 Basisdimensionen<sup>12</sup> Tonhöhe (engl. „pitch“), Dauer (engl. „duration“) und Rhythmus (engl. „rhythm“), Harmonie (engl. „harmony“) und Klangfarbe (engl. „timbre“), zur Beschreibung der Charakteristik eines musikalischen Objektes, z.B. eines Songs.

#### 2.1 Basisdimensionen

##### 2.1.1 Tonhöhe (englisch: pitch)

Die Tonhöhe ist eine musikalische Dimension und kann einzeln bewertet werden. Sie beschreibt die Frequenz eines Tons, einer musikalischen Tonskala oder entsprechende Tonhöhenempfindungen. In einer Notenschrift wird die Tonhöhe anhand der vertikalen Ausrichtung der Note abgelesen. Sollte die Tonhöhe nicht durch eine Notenschrift oder MIDI<sup>13</sup> gegeben sein, muss sie anhand der Wellenform von Audiodateien entnommen werden. Dieser Prozess wird auch als „pitch tracking“ (dt.: Tonhöhenerkennung) bezeichnet.<sup>14</sup>

Eine wichtige Rolle spielt auch die vom Menschen „wahrgenommene Tonhöhe“. Durch sie kann bestimmt werden, wie hoch oder tief ein Ton wahrgenommen wird.

---

<sup>12</sup> Vgl. Stock: Information Retrieval: Informationen suchen und finden (2007), 522-529

<sup>13</sup> Musical Instrument Digital Interface = Datenübertragungs-Protokoll

<sup>14</sup> Vgl. Stock: Information Retrieval: Informationen suchen und finden (2007), 522-529

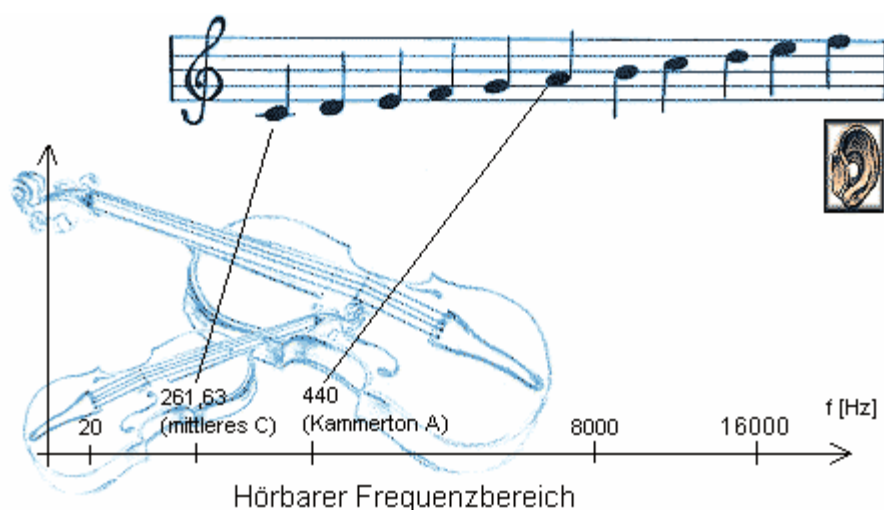


Abbildung 2 : Zusammenhang zwischen Zahlenwert der Frequenzangabe und der empfundenen Tonhöhe

Quelle: [http://www.dasp.uni-wuppertal.de/ars\\_auditus/akustik/akustik21.htm](http://www.dasp.uni-wuppertal.de/ars_auditus/akustik/akustik21.htm)

Abbildung 2 zeigt den Zusammenhang von dem Zahlenwert der Frequenzangaben und der empfundenen Tonhöhe eines Menschen. Durch den Anstieg der Frequenz bzw. Tonhöhe, verändert sich die Position der Note in der Notenschrift. Der „Kammerton A“ liegt bei 440 Hz<sup>15</sup> und legt damit fest, wie ein Instrument gestimmt wird.<sup>16 17 18</sup>

### 2.1.2 Dauer und Rhythmus (englisch: duration and rhythm)

Bei dem Rhythmus handelt es sich um die Zeitstruktur eines Musikstücks, welches die Folge von Notendauern und Pausen beinhaltet. Der Rhythmus ist eine Abfolge von kurzen und langen, von betonten und unbetonten Tönen. Das Grundmaß ist ein durchgehender Schlag, der das Tempo eines Musikstückes angibt. Melodien können bei dieser Dimension weder transformiert noch verzerrt werden, d. h. Komponenten wie Klangfarbe und Tonhöhen können sich ändern, aber der Rhythmus bleibt immer stabil. Somit kann jeder Mensch verstehen, dass es sich um dasselbe Musikstück handelt.

<sup>15</sup> Hertz – Einheit für die Frequenz - Anzahl der Schwingungen pro Sekunde

<sup>16</sup> Vgl. Auswirkung von Amplitude und Frequenz auf die Hörwahrnehmung. [http://www.dasp.uni-wuppertal.de/ars\\_auditus/akustik/akustik21.htm](http://www.dasp.uni-wuppertal.de/ars_auditus/akustik/akustik21.htm) [letzter Zugriff: 12.05.2009]

<sup>17</sup> Vgl. Musiklehre Online – Tonhöhe. [http://www.musica.at/musiklehre/11\\_002.htm](http://www.musica.at/musiklehre/11_002.htm) [letzter Zugriff: 03.05.2009]

<sup>18</sup> Vgl. Inhaltsbasierte Musiksuche: Konzepte und Anwendungen.

[http://www.medien.ifi.lmu.de/fileadmin/mimuc/hs\\_ws0506/papers/MIR.pdf](http://www.medien.ifi.lmu.de/fileadmin/mimuc/hs_ws0506/papers/MIR.pdf) [Letzter Zugriff: 11.06.2009]

Beispiele für Rhythmen sind Marschrhythmus, Walzerrhythmus, Sambarhythmus, Tangorhythmus, etc.,

Im MIR betrachtet, setzt sich die Dimension um die Dauer und den Rhythmus aus dem Tempo, der Länge und der Betonung einer Folge von Noten zusammen.<sup>19 20</sup>

### 2.1.3 Harmonie (englisch: harmony)

Harmonie entsteht bei dem simultanen Erklängen von zwei oder mehrerer „pitches“. Bei diesem Sachverhalt ist von „Polyphony“ die Rede. Das Gegenstück ist die „Monophony“ mit nur einem Ton. Auch Instrumente können daran unterschieden werden. Monophone Instrumente sind z. B. Fagott oder Tuba. Diese können nur einen Ton wiedergeben. Ein polyphones Instrument ist zum Beispiel eine Gitarre.

In der Musiknotation, wird die Harmonie durch verschiedene vertikale Positionierungen von Noten dargestellt. Beispielsweise handelt es sich bei einem Akkord um mindestens drei oder mehrere Noten, die simultan als Hauptklang erklingen.

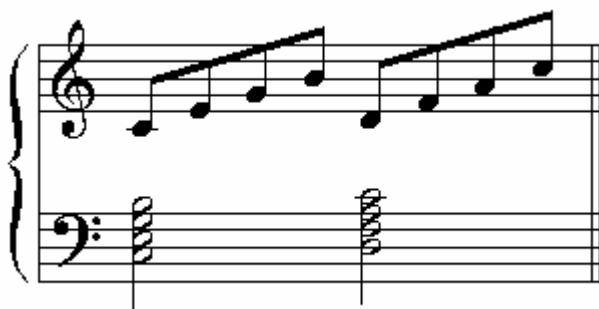


Abbildung 3 : Akkord

Quelle: [www.barpiano-musik.de/html/musik\\_tipps.html](http://www.barpiano-musik.de/html/musik_tipps.html)

Die Aspekte von harmonischen Informationen sind selbst bei in Noten dargestellten Melodien nur schwer zugänglich. Obwohl sie in der Partitur

---

<sup>19</sup> Vgl. Stock: Information Retrieval: Informationen suchen und finden (2007) 522-529

<sup>20</sup> Vgl. Inhaltsbasierte Musiksuche: Konzepte und Anwendungen.

[http://www.medien.ifi.lmu.de/fileadmin/mimuc/hs\\_ws0506/papers/MIR.pdf](http://www.medien.ifi.lmu.de/fileadmin/mimuc/hs_ws0506/papers/MIR.pdf) [Letzter Zugriff: 11.06.2009]

vorhanden sind, finden sie häufig keine explizite Erwähnung in Musikwerken.<sup>21 22</sup>

### 2.1.4 Klangfarbe (englisch : timbre)

Die Klangfarbe ist ein musikalisches Merkmal, welches den Klang zweier Töne, die von der Tonhöhe und Lautstärke identisch sind, unterscheidet. Beispielsweise klingt dieselbe Note auf einer Geige anders als auf einem Piano. Jedes musikalische Instrument besitzt eine eigene Klangfarbe. Sie ist eine Art Gemisch aus Grundton, Obertönen, Rauschanteilen usw., sowie den zeitlichen Verlauf des Frequenzspektrums und der Lautstärke. Besonders wichtig ist das Einschwingverhalten eines Tons, d.h. der zeitliche Verlauf des Spektrums und der Lautstärke in den ersten Sekundenbruchteilen.<sup>23</sup>

Es gibt auch musikalisch genutzte Klänge, die nicht oder nicht allein aus Grundton und Obertönen aufgebaut sind<sup>24</sup>:

- Glockenklänge lassen sich eher durch Grundton und Untertöne beschreiben und ganz generell führt eine nicht-harmonische Überlagerung vieler Sinussignale zu einem eher metallenen Klang
- Trommeltöne sind ein Gemisch aus inharmonischen Schwingungen und Rauschsignalen;
- auch schmalbandiges Rauschen lässt sich als musikalische Klangfarbe nutzen (Windheulen);
- bei vielen musikalischen Klängen sind zudem auch Rauschanteile prägend für die Klangfarbe (z. B. Anblasgeräusche bei Blasinstrumenten und Orgelpfeifen).

---

<sup>21</sup> Vgl. Stock: Information Retrieval: Informationen suchen und finden (2007) 522-529

<sup>22</sup> Vgl. Inhaltsbasierte Musiksuche: Konzepte und Anwendungen.

[http://www.medien.ifi.lmu.de/fileadmin/mimuc/hs\\_ws0506/papers/MIR.pdf](http://www.medien.ifi.lmu.de/fileadmin/mimuc/hs_ws0506/papers/MIR.pdf) [Letzter Zugriff: 11.06.2009]

<sup>23</sup> ebenda

<sup>24</sup> Vgl. Klangfarbe. <http://www.computerbase.de/lexikon/Klangfarbe> [letzter Zugriff: 10.05.2009]



### 2.1.5 Weitere

Weiterhin gibt es noch folgende wichtige Komponenten für MIR:<sup>25</sup>

- Erfassung der gesungenen Texte
- Identifikation einzelner Instrumente
- Erkennung von Sängern

## 2.2 Zusätzliche Begriffe der Musiklehre

In den Bereichen Musiklehre bzw. Musiktheorie existiert ein weites Spektrum an musikalischen Begriffen, die im Rahmen dieser Arbeit nicht alle vorgestellt werden können.

### 2.2.1 Akustik

Die Akustik ist die Lehre vom Schall, dessen Erzeugung und Ausbreitung. Des Weiteren befasst sich die Akustik mit den physiologischen und psychologischen Aspekten des menschlichen Hörens und aus physikalischer Sicht mit den Schallbereichen oberhalb (Ultraschall) und unterhalb (Infraschall) der Hörgrenze. Zusammengefasst liegt der Schwerpunkt auf der Schallerzeugung, Schallaufzeichnung und –messung, sowie der Abstrahlung. Zusätzlich kann durch die Akustik, die Summe der gespielten Klangfarben ermittelt werden. Die Spezialgebiete gliedern sich musikalisch in die Raum- und Bauakustik.<sup>26</sup>

---

<sup>25</sup> Vgl. Stock: Information Retrieval: Informationen suchen und finden (2007), 522-529

<sup>26</sup> Vgl. Algorithmen in Akustik und Computermusik. <http://www.iem.at/lehre/skripten/> [Letzter Zugriff: 11.05.2009]

### 2.2.2 Melodie

Die Melodie ist eine abgeleitete Dimension der 4 Basisdimensionen und ist eine geschlossene Folge von Tönen, die anhand von Intervallen verbunden werden. Diese Verbundenheit beruht auf Dynamik, Tempo, Harmonik und Rhythmik.<sup>27</sup>

### 2.2.3 Takt

Der Takt bezeichnet, im Kontext zum Tempo eines Musikstückes, die äquidistante Unterteilung einer Zeitspanne. Der Rhythmus beruht auf einem Takt, besitzt aber zusätzlich noch Laute, Töne und Schläge. Als zeitliche Maßeinheit dient dem Takt die Zählzeit. Die Definition der Taktart wird bestimmt, durch die Zusammengehörigkeit von Noten eines Notenwertes. Zum Beispiel enthält ein  $\frac{3}{4}$  Takt drei Viertelnoten.<sup>28</sup>

### 2.2.4 Ton

Der Ton ist die kleinste Einheit in der Musik und setzt sich aus folgenden Parametern zusammen: Tonhöhe, Tondauer, Lautstärke und Klangfarbe.<sup>29</sup>

### 2.2.5 Tonleiter und Tonsystem

Eine Tonleiter ist ein stufenweise in bestimmter Intervallfolge angeordneter Ausschnitt, aus dem gesamten Inhalt eines Tonsystems.

Ein Tonsystem ist ein aus einer bestimmten Kultur oder Epoche stammender ausgewählter Vorrat mit der Gesamtheit aller Töne.<sup>30</sup>

---

<sup>27</sup> Vgl. Stock: Information Retrieval: Informationen suchen und finden. (2007) 522-529

<sup>28</sup> Vgl. Holst: Das ABC der Musik: Grundbegriffe, Harmonik, Formen, Instrumente(1992)

<sup>29</sup> ebenda

<sup>30</sup> ebenda

## 2.2.6 Tonalität

Die Tonalität bezeichnet jede Beziehung von Tönen und Akkord, im engeren Sinn das Bezugssystem der Dur- Moll- tonalen Musik.<sup>31</sup>

## 2.2.7 Tonart

Die Aufgabe der Tonart bezieht sich auf die Festlegung des Tongeschlechts (Dur und Moll) auf einer spezifischen Tonstufe. Die Tonart bestimmt für ein Musikwerk einen Grundton mit seinem Dreiklang<sup>32</sup> und innerhalb der Dur- Moll Harmonik ein Bezugssystem aller übrigen Töne und Akkorde.<sup>33</sup>

## 2.2.8 Intervall

Das Intervall bezeichnet den Abstand bzw. das Schwingungsverhältnis von zwei gleichzeitig oder nacheinander erklingenden Tönen.<sup>34</sup>

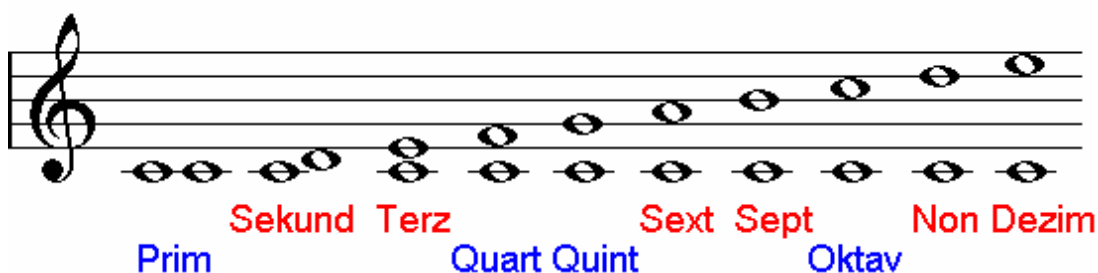


Abbildung 4 : Einfache Intervalle der Musik

Quelle: [http://www.musiklehre.at/7\\_001.htm](http://www.musiklehre.at/7_001.htm)

<sup>31</sup> Vgl. Holst: Das ABC der Musik: Grundbegriffe, Harmonik, Formen, Instrumente(1992)

<sup>32</sup> Erklingen von 3 Tönen gleichzeitig (siehe Akkord)

<sup>33</sup> Vgl. Holst: Das ABC der Musik: Grundbegriffe, Harmonik, Formen, Instrumente(1992)

<sup>34</sup> ebenda

### 2.2.8.1 Oktave und Halbtöne

Die Oktave ist der achte Ton in der diatonischen Tonleiter und das Intervall bei dem die Frequenzen von zwei Tönen im Verhältnis 2:1 stehen. Der Abstand eines Halbtons ist die geringste Entfernung zweier Töne voneinander (das kleinste Intervall). Eine Oktave besteht aus 12 Halbtonsschritten.<sup>35</sup>

### 2.2.9 Grund- und Oberton

In der Harmonielehre wird der Grundton häufig als der tiefste Ton<sup>36</sup> bei einem Akkord bezeichnet. In der Akustik ist er der tiefste Teilton eines Klanges. Weiterhin wird der erste Ton einer Tonleiter als Grundton bezeichnet.

Der Oberton ist ein Vielfaches der Frequenz eines Grundtons. Die Überlagerung von Grund- und Obertönen wird als Klang bezeichnet. Beispielsweise wird der Klang der menschlichen Stimme durch die charakteristische Verteilung der Lautstärke der Obertöne bestimmt. Als weiteres Beispiel dient das Spielen eines Instruments. Die dabei abgegebenen Frequenzen sind ungleichmäßig. Sie erzeugen eine Grundfrequenz, die sich in weitere Frequenzen abspaltet. Zusammengefasst werden diese als Obertöne bezeichnet.<sup>37</sup>

### 2.2.10 Klang

Der Klang wird als Verschmelzung von einem Grundton und den dazugehörigen aufbauenden Teiltönen zu einem Ganzen bezeichnet. Dem daraus entstehenden Schallsignal kann durch das menschliche Gehör eine Tonhöhe zugeordnet werden. Das Spektrum der Teiltöne ist dabei je nach Art des Instruments (oder Vokalstimme) und Dynamik<sup>38</sup> unterschiedlich,

---

<sup>35</sup> Vgl. Holst: Das ABC der Musik: Grundbegriffe, Harmonik, Formen, Instrumente(1992)

<sup>36</sup> tiefste Ton nicht zwangsläufig der Grundton

<sup>37</sup> Vgl. Holst: Das ABC der Musik: Grundbegriffe, Harmonik, Formen, Instrumente(1992)

<sup>38</sup> Tonstärke bzw. Lautstärke

---

sodass verschiedene Klangfarben entstehen. Die Klangsynthese erzeugt und setzt Teiltöne mit elektronischen Mitteln zusammen.<sup>39</sup>

## 2.2.11 Notation

Im Bereich der Musik wird eine Notation als eine „Notenschrift“ bezeichnet. Auf dem 5-zeiligen Notensystem werden die Noten mit ihren Eigenschaften (Tempo, Taktart, Dynamik und Instrumentation) eingetragen. Die Tondauer und die Tonhöhe sind dabei abhängig von der Form und der Position der gesetzten Note.<sup>40</sup>



Abbildung 5 : Notensystem

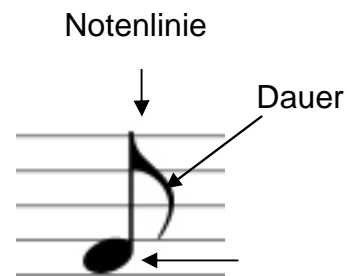


Abbildung 6 : Eigenschaften einer einzelnen Note

Quelle für Abb. 5: [http://www.musicians-place.de/images/harmonielehre\\_01/notensystem-6.gif](http://www.musicians-place.de/images/harmonielehre_01/notensystem-6.gif)

---

<sup>39</sup> Vgl. Holst: Das ABC der Musik: Grundbegriffe, Harmonik, Formen, Instrumente(1992)

<sup>40</sup> ebenda

## Kapitel 3 : Repräsentation von Musik

### 3. Musikrepräsentation

Das Kapitel „Musikrepräsentation“ befasst sich mit den wichtigsten Formaten bzw. Codes des MIR zur Speicherung, Aufzeichnung und Wiedergabe von Signalen. Musik kann auf unterschiedliche Weisen dargestellt sein. Rein technisch werden diese in 2 Kategorien unterschieden (siehe Abbildung 7).

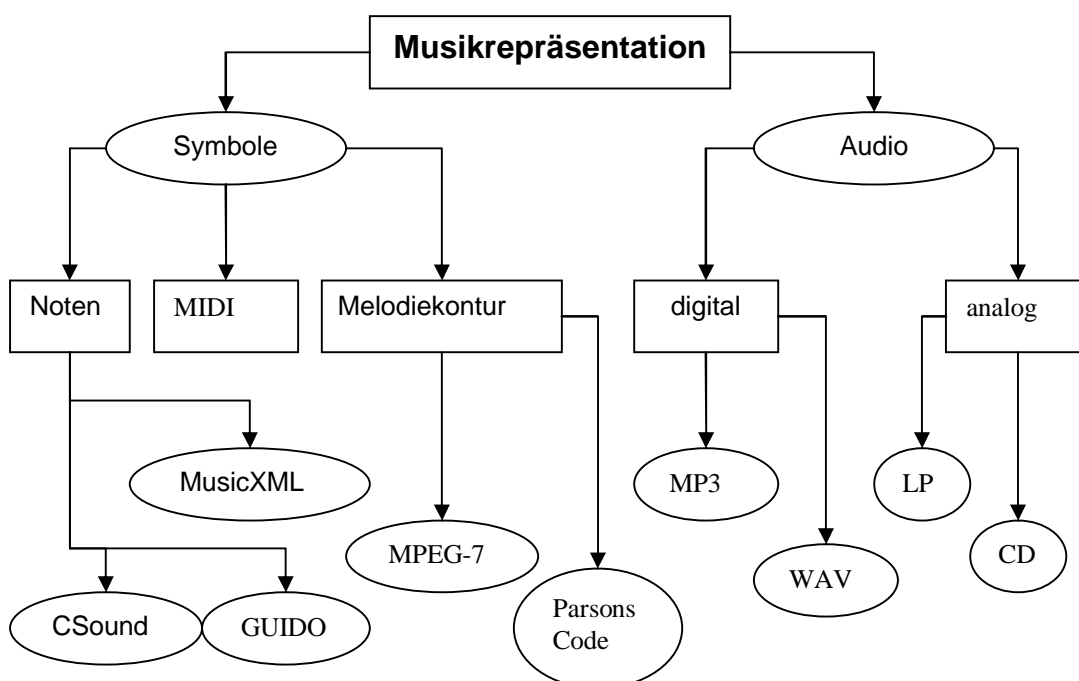


Abbildung 7 : grobe Übersicht zur Musikrepräsentation<sup>41</sup>

Music format	Example	Compared to image retrieval	Compared to text retrieval	Structure	Convert to lower format	Convert to higher format
music notation (Finale, Sibelius, MusicXML)		compound objects	text + markup	much	easy (OK job)	--
time-stamped events (MIDI)		objects, scenes	text	little	easy	fairly hard (OK job)
digital audio (MP3, Wav)		primitive features	speech	none	--	hard

Abbildung 8 : Darstellung von Musikformaten

Quelle: <http://www.cs.uu.nl/groups/MG/multimedia/publications/art/mme-music.pdf>

<sup>41</sup> Vgl. Batke (2006): Untersuchung von Melodiesuchsystemen sowie von Verfahren zu ihrer Funktionsprüfung, S. 25

### 3.1 Audio-Repräsentation

Audioaufnahmen werden in einer digital codierten Wellenform (Waveform) bzw. Luftdruckschwingung dargestellt. Diese ist die Basis von CDs mit Audioaufnahmen und hängt mit den physikalischen Ursachen des Schalls zusammen. Dieser wird von vibrierenden Objekten erzeugt und die entstehenden Schwingungen auf ein Trägermedium übertragen. Das Medium wird dabei durch die Schwingung gestaucht und gedehnt, was zu Veränderungen des Drucks führt. Wenn dieser Druck nun in der Nähe des schwingenden Objekts in einem Zeit/Luftdruck Graphen aufgezeichnet wird, erhält man die Darstellung in Form einer Welle. Die einfachste Erscheinungsform von Wellen ist eine Schwingung, die dem Verlauf einer Sinuskurve entspricht.<sup>42</sup>

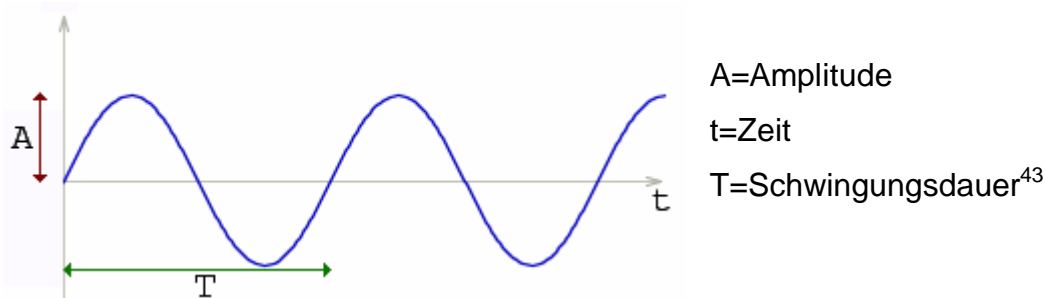


Abbildung 9 : harmonische Sinusschwingung

Quelle: [http://wiki.delphigl.com/images/4/47/Tutorial\\_SoftSynth\\_welle.png](http://wiki.delphigl.com/images/4/47/Tutorial_SoftSynth_welle.png)

Audiosignale sind Schallwellen und nur sehr selten eine reine Sinusschwingung. Sie bestehen aus einer Überlagerung von theoretisch unendlich vielen Sinuskurven und sind somit nicht periodisch. Es gibt 2 Größen die Schallwellen bestimmen: Amplitude und Frequenz. Die Amplitude beschreibt bei einer Welle die maximale Auslenkung aus der Mittellage zu einem bestimmten Zeitraum. Sie bestimmt die übertragene Energie der Welle und somit die wahrgenommene Lautstärke. Die Frequenz beschreibt die Häufigkeit einer Schwingung in einem festgelegten Zeitraum und bestimmt die wahrgenommene Tonhöhe. Jede der theoretisch unendlich vielen Sinusschwingungen besitzt eine eigene Frequenz und Amplitude.

<sup>42</sup> Vgl. Effiziente Methoden zur hochauflösenden Musiksynchronisation. [http://www.mpi-inf.mpg.de/~mmueller/publications/2008\\_Diplomarbeit\\_EwertSebastian\\_MusikSync.pdf](http://www.mpi-inf.mpg.de/~mmueller/publications/2008_Diplomarbeit_EwertSebastian_MusikSync.pdf) [letzter Zugriff: 12.06.2009]

<sup>43</sup> bzw. Periodendauer, Dauer einer vollständigen Schwingung der Welle

Zusammengefasst entsteht eine normale Schallwelle aus der Addition der mathematischen Überlagerung aller unendlich vielen Sinuskurven.

Die *analoge Übertragung* dient der Übermittlung von Audiosignalen. Der Begriff „analog“ bedeutet gleichartig und bezeichnet Vorgänge, die unterschiedlich sind, aber denselben Gesetzmäßigkeiten unterliegen. Schallwellen und elektromagnetische Wellen verhalten sich analog zueinander. Bei der Überführung von Schall in elektromagnetische Wellen ändert sich die Beschaffenheit nicht.<sup>44</sup>

### 3.1.1 Abtastung (englisch: Sampling)

Die Abtastung ist eine Voraussetzung, um ein analoges Signal in ein Digitalsignal umzuwandeln. Das Audiosignal wird bei der Aufzeichnung in bestimmten Zeitabständen gemessen bzw. „abgetastet“. Die Schallwellen werden in elektromagnetische Schwingungen übertragen, d.h. der Strompegel ändert sich fortwährend und wird gemessen. Dieser Prozess wird als „Abtasten“ bezeichnet und die entstandenen Abtastwerte sind „Samples“. Während der Abtastung gehen Daten verloren. Der Verlust ist aber kaum hörbar so lange die Abtastrate hoch genug ist. Beispielsweise bei der Aufnahme in CD-Qualität ist eine Messung des Amplitudenwertes 44100 mal in der Sekunde nötig, dabei entspricht die Abtastrate 44,1 Kilohertz. Zum Vergleich nimmt das menschliche Gehör einen Frequenzbereich von 20Hz bis 20.000 Hz wahr.<sup>45 46</sup>

---

<sup>44</sup> Vgl. Audiosignale. <http://www.itec.uka.de/seminare/rftk/mp3/#ToC7>

<sup>45</sup> Vgl. Plato (2004): Übungsbuch zur Numerischen Mathematik: Aufgaben, Lösungen und Anwendungen, S. 15

<sup>46</sup> Vgl. Audiosignale. <http://www.itec.uka.de/seminare/rftk/mp3/#ToC7>



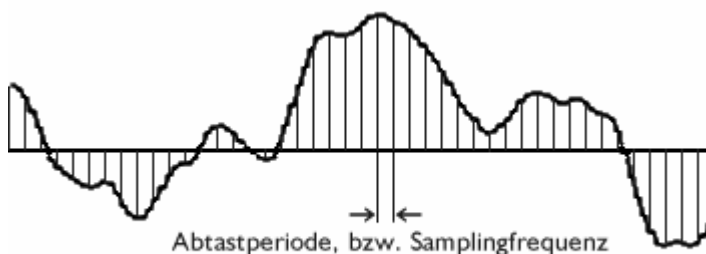


Abbildung 10 : Abtastung eines Audiosignals

Quelle: <http://www.itec.uka.de/seminare/rftk/mp3/>

### 3.1.2 Quantisierung

Der Verlust bei der Abtastung kann anhand der höheren Abtastrate gemildert werden. Dennoch ist die Digitalisierung verlustbehaftet, denn ein Computer kann nur ganze bzw. endliche Signale verarbeiten. Die vorher abgetasteten Pegelwerte sind das Gegenteil. Sie können jeden Wert zwischen Ober- und Untergrenze annehmen. Aus diesem Grund werden die Werte gerundet. Dieser Vorgang wird als „*Quantisierung*“ bezeichnet. Jeder Abtastwert wird als Ganzzahl mit einer bestimmten Auflösung in Bit gespeichert. Die höchste Zahl in dieser Auflösung entspricht der Obergrenze der Pegelausschläge und die kleinste der Untergrenze. Dabei zählt je höher die Auflösung, umso mehr Zwischenwerte zwischen Ober- und Untergrenze unterschieden werden können, desto präziser wird das Signal gespeichert. Zusammengefasst bedeutet die Quantisierung, die Speicherung eines Wertes mit bestimmter Auflösung.<sup>47</sup>

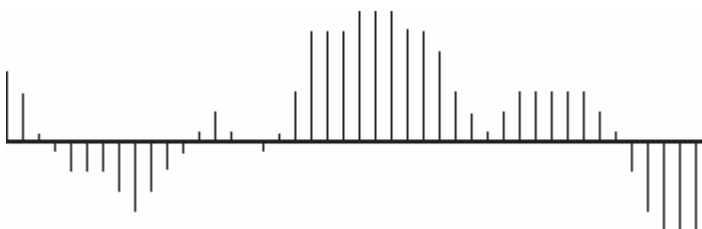


Abbildung 11 : Darstellung<sup>48</sup> der Quantisierung

Quelle: <http://www.itec.uka.de/seminare/rftk/mp3/#ToC7>

<sup>47</sup> Vgl. Audiosignale. <http://www.itec.uka.de/seminare/rftk/mp3/#ToC7>

<sup>48</sup> je mehr Rundungen, desto ungenauer das Signal – niedrige Auflösung

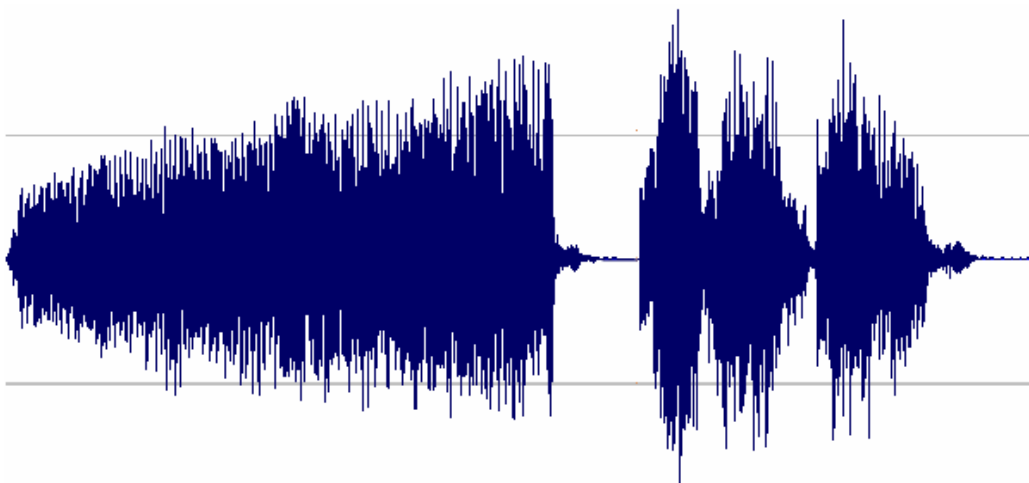


Abbildung 12 : Darstellung einer Wellenform

Quelle: [http://www.sputnik.de/sixcms\\_upload/media\\_fast/43/T5\\_wellenform.jpg](http://www.sputnik.de/sixcms_upload/media_fast/43/T5_wellenform.jpg)

### 3.1.3 PCM : Puls-Code-Modulation

PCM ist ein Modulationsverfahren mit dem Ziel der digitalen Übertragung von analogen Signalen. Die Signale von PCM lassen sich leicht speichern, verarbeiten und übertragen. Ursprünglich war es für Telefonanlagen gedacht, mittlerweile sind es wesentlich mehr Einsatzgebiete (Audio-, Digital-Video- und CD-Formate, Telemetrie und virtuelle Realität), in denen es genutzt wird. Insgesamt besteht das gesamte Verfahren aus mehreren Schritten<sup>49</sup>: Sample and Hold, Quantisierung und Codierung. Das analoge Signal wird in gleichmäßigen Abständen abgetastet und gespeichert. Das Abtastspektrum wird in Stufen eingeteilt. Dabei gilt, je mehr Stufen bei der Digitalisierung gewählt werden, desto genauer wird das Signal verarbeitet. Die Anzahl der Stufen entspricht der Auflösung und wird in Bit angegeben.<sup>50</sup>

24 Bit (16,7 Mio. Stufen)	sehr hohe Qualität, Studioeinsatz, geplanter DVD-Standard
16 Bit (65.536 Stufen)	hohe Qualität, Audio-CDs, Multimedia-Präsentationen
8 Bit (256 Stufen)	niedere Qualität, nicht zu empfehlen

---

<sup>49</sup> siehe Anhang

<sup>50</sup> Vgl. Digitale Audioaufnahme nach dem PCM-Verfahren.

<http://wiki.arch.ethz.ch/twiki/pub/Extern/FormatWav/Audioinfoblatt.pdf> [Letzter Zugriff: 13.06.2009]

Aufgrund der Stufenbildung entsteht eine Ungenauigkeit, welche als „*Quantisierungsfehler*“ bezeichnet wird und als Rauschen wahrgenommen werden kann. Aus diesem Grund wird das „*Oversampling*“ (*Übertastung*) verwendet.<sup>51</sup>

### 3.1.4 WAV oder WAVE

Das Wave Dateiformat ist ein Containerformat für die Speicherung von Audiodaten und demzufolge kein Audioformat. Es basiert auf RIFF (Resource Interchange File Format), das von Microsoft und IBM für das Betriebssystem Windows entwickelt wurde. In WAV sind häufig PCM Daten enthalten.

Im Gegensatz zu MP3 sind WAV Dateien unkomprimiert und somit verlustfrei, leider aber auch dementsprechend groß. Es erfolgt keine Datenkompression, stattdessen werden nur die Rohdaten gespeichert. Das Format besitzt 3 Datenblöcke, so genannte „*chunks*“, die verschiedene Aufgaben erfüllen. Das „*RIFF-chunk*“ identifiziert die Datei als WAV Datei. Das „*FORMAT-chunk*“ beinhaltet Informationen über die Anzahl der Kanäle (mono oder stereo), Samplerate und weitere Parameter. Weiterhin sind die Sample-Rate<sup>52</sup> und die Bits pro Sample<sup>53</sup> enthalten. Der dritte Block („*DATA-chunk*“) enthält die eigentlichen Audiodaten.<sup>54</sup>

### 3.1.5 MP3

MP3 ist ein sehr populäres Dateiformat für Musik. Es handelt sich um eine verlustbehaftete Kompression basierend auf MPEG-1-Audio Layer-3. Dabei besitzt es eine hohe Wiedergabequalität bei einem Kompressionsfaktor von über 10:1. MP3 ist ein verlustbehaftetes Format, da bei der Kompression für Menschen nicht hörbare Signalanteile herausgefiltert werden. Das

---

<sup>51</sup> Vgl. PCM – Pulsmodulation. <http://www.elektronik-kompodium.de/sites/kom/0312281.htm> [Letzter Zugriff: 15.06.2009]

<sup>52</sup> Abtastrate des Klages

<sup>53</sup> Wertebereich, den die Amplitude annehmen kann

<sup>54</sup> Vgl. WAV-Format. [http://swlab.et.fh-duesseldorf.de/pc\\_pool/lernmodule/multimediateien/Kapitel33.htm](http://swlab.et.fh-duesseldorf.de/pc_pool/lernmodule/multimediateien/Kapitel33.htm) [letzter Zugriff: 15.05.2009]

menschliche Ohr kann die Signale nicht wahrnehmen, jedoch bei einer niedrigen Datenrate wird die Datenreduktion hörbar. Zur besseren Identifikation, gibt es MP3-Dateien mit *ID-3* Tags. Diese bieten die Möglichkeit, Metadaten über das Stück wie z.B. Titel, Interpret und Album zu speichern. Die Informationen werden dabei manuell eingegeben und müssen einheitlich sein. Vergleichsweise besitzt ein normaler Standard Popsong ca. 30-35 Megabyte. Im MP3 Format benötigt dieser drei bis vier Megabyte oder weniger mit Qualitätsverlust. Das Format unterstützt 3 Hauptmodi: Mono, Stereo und Join Stereo. Neben dem normalen Standard MP3 Format, wurde das Enhanced MP3<sup>55</sup> entwickelt, welches Zusatzinformationen integrieren kann. Der aktuelle Stand ist die Entwicklung von mp3PRO, welches hohe Frequenzen auch bei einer niedrigen Bitrate ermöglicht.<sup>56 57</sup>

### 3.1.6 Weitere

Es existiert eine Vielzahl von Audioformaten, die aber nicht alle für den MIR Bereich ausschlaggebend sind, bzw. Formate die mehr oder weniger verwendet werden. Einige Beispiele

verlustfrei		verlustbehaftet
unkomprimiert	komprimiert	ogg : Ogg-Vorbis
CD-DA : Compact Disc Digital Audio	flac : Free Lossless Audio Compression	
voc : Creative Labs Sound-Blaster-Sprach- Dateiformat	ALAC : Apple Lossless Audio Codec	mp2 : MPEG 1 Layer 2
AIFF: Audio Interchange File Format (Containerformat)	adpcm : Adaptive- Differential-Pulse-Code- Modulation	mp1 : MPEG 1 Layer 1

Tabelle 1 : Liste von diversen Audioformaten

<sup>55</sup> ergänztes MP3 Format, visuelle Informationen zum Titel sind sichtbar (Liedtexte, Konzertdaten usw.)

<sup>56</sup> Vgl. Music Information Retrieval Automatische Genre-Klassifikation. <http://cvpr.uni-muenster.de/teaching/ss05/blockseminarSS05/downloads/Music-Ausarbeitung.pdf> [letzter Zugriff: 20.05.2009]

<sup>57</sup> Vgl. mp3PRO. <http://www.itwissen.info/definition/lexikon/mp3Pro-mp3Pro.html>

### 3.2 Symbolische Repräsentation

Die zweite Darstellungsform beruht auf „Symbolen“, beispielsweise Noten einer Partitur<sup>58</sup>. Diese Form ist weniger physikalisch geprägt als ihr Pendant, die digitalisierten Schallwellen. Eine Notenschrift zeigt dem Musiker, wie ein Instrument gespielt werden muss, in Bezug auf Rhythmus, Tonhöhe etc.. Diese Informationen sind anhand von Noten in einem Notensystem sichtbar. Dabei handelt es sich um grafische Zeichen zur Aufzeichnung eines Tones oder Schlages. Die Platzierung in einem Notensystem (bzw. Notenlinien) zeigt die Tonhöhe der jeweiligen Note an. Die äußere Form widerspiegelt die Tondauer bzw. den Notenwert. Menschen nehmen Frequenzen nicht linear wahr, sondern logarithmisch, was dazu führt, dass 880Hz nur dreimal so hoch wie 220Hz empfunden wird. Für den Computer müssen technische Darstellungsformen produziert werden bzw. besondere Formate die Noten elektronisch speichern und für den Einsatz in Notationssystemen nutzbar machen. Eine Art der symbolischen Darstellung ist die „piano-roll“ für MIDI.<sup>59</sup>

60

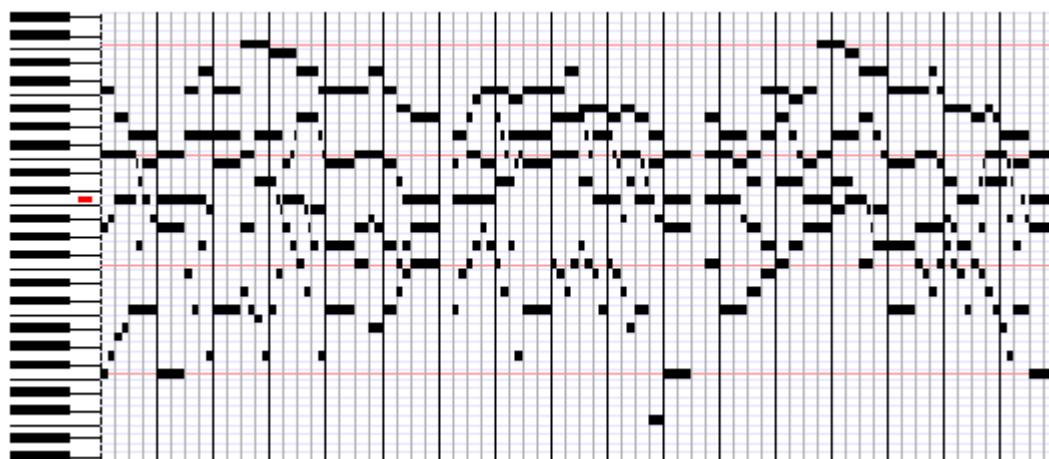


Abbildung 13 : piano-roll

Quelle: <http://www.umsl.edu/~fraundorfp/musictalkcrystals.html>

<sup>58</sup> angeordnete Zusammenstellung aller Einzelstimmen einer Komposition

<sup>59</sup> Vgl. Music Information Retrieval Automatische Genre-Klassifikation. <http://cvpr.uni-muenster.de/teaching/ss05/blockseminarSS05/downloads/Music-Ausarbeitung.pdf> [letzter Zugriff: 20.05.2009]

<sup>60</sup> Vgl. What is MIDI?. <http://www.midi-classics.com/whatmidi.htm> [letzter Zugriff: 20.05.2009]

### 3.2.1 MIDI

MIDI ist eines der primären Formate im Bereich MIR. Es wurde 1981 von Dave Smith für die „*Audio Engineering Society*“ entwickelt. Es steht als Abkürzung für „*Musical Instrument Digital Interface*“ und ist ein Protokoll zur digitalen Datenkommunikation. Die Schnittstelle ermöglicht eine Kommunikation zwischen Instrumenten, Synthesizern, Effektgeräten, Mischpulten und Computern. MIDI ist ausschließlich dazu in der Lage musikalische Steuerinformationen aufzuzeichnen, wiederzugeben und zu speichern, jedoch keine Audiosignale. Hauptsächlich werden Ereignisse beschrieben im Sinne von „Note-an“ und „Note-aus“, der Wechsel von Voreinstellungen für Instrumentenklänge, die Betätigung des Haltepedals, gleitende Änderungen der Tonhöhe und Zeitinformationen. Durch die Erweiterung kommen zusätzlich Faktoren hinzu, wie Zeitstempel oder Zeitmarken, Klangmuster, sowie die Dateiformate MIDI Standard<sup>61</sup> und General MIDI Standard. Letzterer ist 24stimmig und besitzt 128 Klänge, mit den dazu beigeordneten Programmwechselnummern, die das Ansprechen von 16 MIDI- Kanälen ermöglichen. Der erste General MIDI Standard besitzt noch den Zusatz „Level 1“. Das Nachfolgemodell „Level 2“ besitzt die doppelte Anzahl an Klängen, insgesamt 256. Die Polyphonie ist auf 32 Stimmen angestiegen.

Das Protokoll verwendet kurze Bytefolgen zum Austausch von Signalen. Die Aufzeichnung, Bearbeitung und Wiedergabe als Musikstück erfolgt dann durch einen Sequenzer<sup>62</sup>.

### 3.2.2 NIFF und MusicXML

Das NIFF Format wird als das erste Standardformat zur Speicherung von Musiknotation<sup>63</sup> bezeichnet und steht als Abkürzung für „*Notation Interchange File Format*“. Das Format besitzt eine hierarchische Struktur,

---

<sup>61</sup> 1990 Standard-MIDI-File-Format (SMF): gespeicherte Musikstücke können auf verschiedene Computertypen geladen und in Musiksoftware bearbeitet werden

<sup>62</sup> Vgl. MIDI Kompendium. [http://www.zem-college.de/midi/mc\\_allg.htm](http://www.zem-college.de/midi/mc_allg.htm) [letzter Zugriff: 16.05.2009]

<sup>63</sup> weitere Dateiformate zur Musiknotation <http://www.music-notation.info/de/compmus/notationformats.html>

ähnlich XML, obwohl es binär ist. Abgelöst wurde NIFF durch das Format MusicXML.

MusicXML dient dem Austausch von Partituren. Es enthält Noteninformationen, sowie (im Gegensatz zu MIDI) auch Layoutinformationen. Weiterhin ist es ein sehr beliebtes Format, da sich die enthaltenen Daten mit sämtlichen, wichtigen Notationsprogrammen austauschen lassen.<sup>64 65</sup>

```
<score-partwise version="2.0">
<work>
  <work-number>Op. 98</work-number>
  <work-title>An die ferne Geliebte (Page 1)</work-title>
</work>
<identification>
  <creator type="composer">Ludwig van Beethoven</creator>
  <creator type="lyricist">Aloys Jeitteles</creator>
  <rights>Copyright © 2002 Recordare LLC</rights>
```

Abbildung 14 : Ausschnitt eines MusicXML Samples

Quelle: <http://www.recordare.com/xml/samples.html>

### 3.2.3 CSound

Bei „CSound“ handelt es sich um einen Software Synthesizer bzw. eine Programmiersprache, welche die Möglichkeit bietet, einen virtuellen Synthesizer selbst zu programmieren. Spezialisiert ist es auf Algorithmen für die Klangsynthese. Die Sprache bietet wesentlich mehr Ausdrucksmöglichkeiten als MIDI und ist somit für Komponisten sehr gut geeignet, da es über eine eigene Partitur -Sprache verfügt. Dabei ist es notwendig, Grundlagenwissen im Bereich digitaler Signalverarbeitung zu besitzen. Das Ziel von CSound ist die Erzeugung einer Audiodatei. Dies geschieht durch die Verarbeitung zweier Quelldateien, einer Orchestra (Instrumente)-Datei und einer Score (Partitur)-Datei. Eine zweite Möglichkeit

---

<sup>64</sup> Vgl. Codierungsformen von Musik.  
[http://www.adwmainz.de/fileadmin/adwmainz/MuKo\\_Veranstaltungen/S2-Digitale\\_Medien/kepper.pdf](http://www.adwmainz.de/fileadmin/adwmainz/MuKo_Veranstaltungen/S2-Digitale_Medien/kepper.pdf)  
[letzter Zugriff: 20.05.2009]

<sup>65</sup> Vgl. MusicXml Definition. <http://www.recordare.com/xml.html> [letzter Zugriff: 21.05.2009]

wäre das Nutzen einer XML Datei, die Angaben für Instrumente und Partitur enthält.<sup>66</sup>

### 3.2.4 Common Music Notations

*Common Music Notation* ist eine freie Software zur Musiknotation. Es wurde 1990 von Heinrich Taube entwickelt und unterstützt zahlreiche Formate. CMN ist geschrieben in der Programmiersprache „Common Lisp“ und läuft ohne Probleme auf einer Vielzahl von Betriebssystemen.

Eingabe<sup>67</sup>:            (cmn staff treble c4 q)

Ausgabe<sup>68</sup>:



### 3.2.5 Parsons Code

Benannt wurde der Parsons Code nach Denys Parsons. Die Hauptaufgabe besteht darin, Melodien zu klassifizieren und zu finden. Das geschieht anhand der Melodiekontur. Sie gibt an, ob sich die Tonhöhe einer Melodie nach oben oder unten verändert oder gleich bleibt. Dazu werden 4 Symbole genutzt, mit denen die Kontur beschrieben wird (siehe Tabelle 2).<sup>69</sup>

Symbol	Bedeutung
d	= down (Tonhöhe sinkt)
u	= up (Tonhöhe steigt)
r	= repeat (Tonhöhe bleibt gleich)
*	= erster Ton als Referenzton

Tabelle 2: Symbole des Parsons Code

<sup>66</sup> Vgl. What is CSound?. <http://www.csounds.com/whatis/index.html> [letzter Zugriff: 20.05.2009]

<sup>67</sup> Vgl. Common Music Notations. <http://ccrma.stanford.edu/software/cmnm/cmnm/cmnm.html> [letzter Zugriff: 22.05.2009]

<sup>68</sup> ebenda

<sup>69</sup> Vgl. Batke (2006): Untersuchung von Melodiesuchsystemen sowie von Verfahren zu ihrer Funktionsprüfung, S. 27



## Kapitel 4 : Audiostandards

### 4. Standards für den Audio Bereich

Eine standardisierte Darstellung ist gerade im Bereich der digitalen Musik, insbesondere dem Segment der Online-Musik, von essentieller Bedeutung. Darum ist es gerade herausfordernd, Mechanismen zu entwickeln, die ein automatisiertes Beschreiben und eine standardisierte Verwaltung für Anbieter als auch Nutzer gewährleisten können. Besonders für Nutzer sollte die Möglichkeit eines schnellen und effizienten Zugriffs auf Informationen in Datenbanken gegeben sein. Einige der heutigen Metadatenstandards eignen sich sehr gut zur Beschreibung von Musikinformation, wobei im Verlauf der Arbeit ausschließlich auf die Standards der Moving Pictures Expert Group eingegangen wird. Weiterhin gibt es SMIL, EBU, SMPTE, TV-Anytime, DIG-35 OCLC/RLG oder Dublin Core usw..

#### 4. 1 Moving Pictures Expert Group

Die Moving Pictures Expert Group wurde 1989 gegründet und ist eine Arbeitsgruppe der ISO (*International Organization for Standardization*). Die Hauptaufgabe der Gruppe besteht in der Entwicklung von Komprimierungs- und Verarbeitungsverfahren für Audio-, Video- und Bilddaten, sowie deren Standardisierung. Tabelle 3 zeigt einen kurzen Überblick, über die bisher erschienenen Standards und deren Inhalte. MPEG-7 und MPEG-21 werden im Verlauf näher erläutert.<sup>70</sup>

Standard	Bedeutung
MPEG-1	Standard zur verlustbehafteten Video- und Audiodatenkompression - Speicherung von Audio- und Videodaten auf einer CD

<sup>70</sup> Vgl. Bröder, Aline(2005); Standards zur Modellierung von Metadaten für digitale Musik. [http://is-frankfurt.de/veranstaltung/Groffmann\\_SS05/Thema%205\\_Broeder\\_dl.pdf](http://is-frankfurt.de/veranstaltung/Groffmann_SS05/Thema%205_Broeder_dl.pdf) [letzter Zugriff: 23.06.2009]

MPEG-2	generischer MPEG-Standard zur Videodekodierung mit Videokompression und Audiokodierung mit Audiokompression
MPEG-3	Erschien leider nie, da die Erweiterung von MPEG-2 ausreichte
MPEG-4	Beinhaltet Beschreibungen von Verfahren zur Video- und Audiodatenkompression

Tabelle 3 : Überblick von MPEG-Standards

#### 4.1.1 MPEG-7

Der MPEG-7 Standard ist für Multimedia und Informationssysteme konzipiert worden. Er unterscheidet sich stark von den früheren Standards, was die heute gängigen Suchmaschinen zu verantworten haben. Als der MPEG-1 Standard im Jahre 1993 entwickelt wurde, war die Technik des Internet noch nicht ausgereift. Die Standards 1-4 befassten sich mit der Beschreibung bzw. Codierung der Daten, wobei MPEG-7 den Fokus auf die Beschreibung von AV-Inhalten setzte. MPEG-7 ist kein Verfahren zur audiovisuellen Kompression im Gegensatz zu den früheren MPEG-Standards. Die Grundidee ist, den audiovisuellen Dateien Beschreibungen beizulegen, um eine bessere Erschließung zu ermöglichen. Im Gegensatz zu Dublin Core, welches einen vordefinierten Katalog an Metadatenfeldern nutzt, werden in MPEG-7 analyseabhängige Beschreibungen von Werten und Inhalten geliefert. Diese werden semantisch und strukturell aufgearbeitet.<sup>71</sup>

#### Einige Anwendungsgebiete

- Entertainment
- Journalismus
- Multimedia-Datenbanken

---

<sup>71</sup> Vgl. Verwendung des MPEG-7 Standards in der Szenen Interpretation. <http://www.sts.tu-harburg.de/pw-and-m-theses/2005/bort05.pdf> [letzter Zugriff: 22.06.2009]

- Informationssysteme
- ...

Der MPEG-7 Standard setzt sich aus 3 Hauptbestandteilen und 7 Teilbereichen zusammen.

## Hauptbestandteile

### Descriptor

Das Element *Descriptors* enthält Metadaten, die den Inhalt kennzeichnen. Dieser besteht beispielsweise aus: Titel, Autor, Copyright-Information etc.. Zur Erstellung eines Descriptors werden die Beschreibungsinformationen extrahiert. Das geschieht manuell oder automatisch. Ein Descriptor enthält mehrere Werte einer Eigenschaft und diese können schwanken zwischen einfachen textbasierten Anmerkungen, bis hin zu komplexen Signalverläufen.<sup>72</sup>

### Descriptor Scheme

Das *Descriptor Scheme* repräsentiert die Struktur und den Zusammenhang zwischen den verschiedenen *Descriptors* und diversen *Description Schemes*.

<sup>73</sup>

### Deskriptor Definiton Language

DDL ist eine standardisierte Sprache auf Basis von XML-Schema, die dem Nutzer eine Erstellung von *Descriptors* und *Description Schemes* ermöglicht. Sie ist plattform- und applikationsunabhängig.<sup>74</sup>

---

<sup>72</sup> Vgl. Bortey, Daniel (2005): Die Verwendung des MPEG-7 Standards in der Szenen Interpretation. <http://www.sts.tu-harburg.de/pw-and-m-theses/2005/bort05.pdf> [letzter Zugriff: 22.06.2009]

<sup>73</sup> ebenda

<sup>74</sup> ebenda

## Die 7 Teilbereiche des Standards<sup>75</sup>

1. Systems
2. Data Definition Language
3. Visual Descriptors
4. Audio Descriptors
5. Multimedia Descriptor Schemes
6. MPEG-7 Reference Software
7. Conformance Testing

## Rolle des MPEG-7 im Audiobereich

Es gibt speziell definierte *Descriptors* und *Descriptor Schemes*, die zur Erkennung von Sprache und einer robusten Identifizierung von Musikstücken genutzt werden. Insbesondere die *Low Level* Merkmale beschreiben und erfassen Spektrum und Cepstrum<sup>76</sup>, Wellenform, Klangfarbe, Grundfrequenz und Obertöne, Signalparameter usw.. In MPEG-7 ist ein Framework für die automatische Erkennung von Audiodateien verankert, aufgrund dessen soll eine eindeutig identifizierbare Beschreibung zum Wiedererkennen von Audiomaterial gewährleistet sein. Zum Beispiel durch *Melody description* und *audio matching*.

## Melody Description und Audio Matching

Die *Melody Description* dient dem Zweck, bei einer fehlerhaften oder unvollständigen Eingabe (z.B. falsche Tonart oder Tempo), dennoch ein korrektes Ergebnis zu finden. Die *Description* muss auf eine robuste und flexible Art die Eigenschaften der Melodie des Stückes durch eine Repräsentation einer Folge der relativen Tonhöhen speichern. Mit einer Werteskala (5-stufig: -2, -1, 0, +1, +2) für die Tondifferenz und das Abspeichern eines quantisierten Rhythmus ist eine kompakte

---

<sup>75</sup> siehe Anhang

<sup>76</sup> Spektrum des logarithmierten Frequenzspektrum

Beschreibungsart der Melodie möglich. Die Erzeugung einer solchen Beschreibung ist beispielsweise aus einer MIDI Datei möglich.

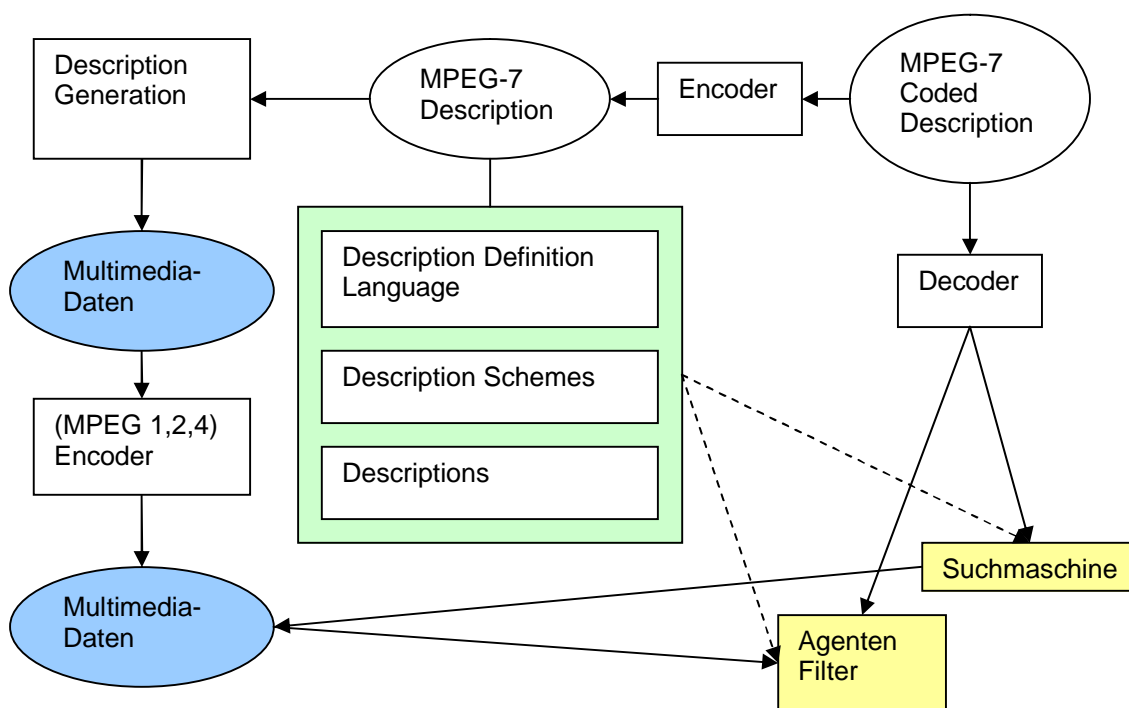


Abbildung 15 : tragende Rolle des MPEG-7 Standard in der Musikverarbeitung des Internet<sup>77</sup>

Das *Audio Matching* ermöglicht eine inhaltsbasierte Erkennung von Audiodateien, anhand eines Vergleichs mit Referenzdaten. Dabei ist die Robustheit von Bedeutung. Die Erkennung soll selbst bei geschnittenen Dateien möglich sein.<sup>78</sup>

Anwendungsbeispiele:

- gezielte Suche nach bestimmten Musikstücken
- Suche nach ähnlichen Stücken
- automatische Zuordnung von Metadaten

<sup>77</sup> Vgl. Ilic, Manoela(2004): MUSIK-METADATEN. <http://www-lehre.inf.uos.de/~milic/Coxi/Musik-Metadaten.pdf> [letzter Zugriff: 23.06.2009]

<sup>78</sup> Vgl. MPEG-7. <http://tams-www.informatik.uni-hamburg.de/lectures/2001ws/vorlesung/medientechnik/18-mpeg7.pdf> [letzter Zugriff: 23.05.2009]

### 4.1.2 MPEG-21

MPEG-21 für Multimedia Framework, dient der Speicherung und Beschreibung diverser Multimediainhalte. Ziel des Standards ist die Nutzung von *Multimedia Ressources* für eine Vielzahl von Netzwerken und Geräten. Zusätzlich wird das Konzept „*Digital Item (ID)*“ eingeführt. ID ist ein eindeutig identifizierbares, digitales Objekt, welches durch eine standardisierte Repräsentation, Identifikation und Metadaten definiert ist. Es besteht aus Mediendaten bzw. Ressourcen und den zugehörigen Metadaten in einer eingebetteten Struktur. Die Ressourcen sind die Deskriptoren von MPEG-7. Der MPEG-21 Standard ist für *Query by Humming* Systeme anwendbar. Die Ergebnisse in der Suchanfrage, die aus Audiodateien, Lied und Notentext, Künstlerinformationen, Informationen zum Album etc. bestehen, könnten mit einem ID beschrieben werden. Zusammengefasst stellt MPEG-21 eine Infrastruktur zur elektronischen Konzeption, Produktion, Freigabe, Vermarktung und Handel von multimedialen Inhalten dar.<sup>79</sup>

---

<sup>79</sup> Vgl. Batke, Johann M.: Untersuchung von Melodiesuchsystemen sowie von Verfahren zu ihrer Funktionsprüfung, S. 62-63

## Kapitel 5 : Merkmalsextraktion

### 5. Grundlagen der Merkmalsextraktion (englisch : Feature extraction)

Die Merkmalsextraktion beschreibt die Gewinnung von wesentlichen Daten aus einem Audiosignal, welche an die menschliche Unterscheidungsfähigkeit angelehnt ist. Zunächst ist es dabei von Bedeutung, dass die Audio-Signale durch Berechnungsverfahren in digital abgetasteter Form vorliegen. Dabei lassen sich die klanglichen Eigenschaften mit Frequenzspektren zeitlich begrenzter Abschnitte (Intervalle) beschreiben. Die Daten können daraufhin anhand verschiedener Verfahren der Signalverarbeitung extrahiert und in datenreduzierter Form gespeichert und weiterverarbeitet werden. Die charakteristischen Eigenschaften werden als *Merkmale* bezeichnet. Beispielsweise in der Akustik ist bekannt, dass die Tonhöhe anhand der Grundfrequenz festgelegt werden kann. Aus diesem Grund wird die Wellenform eines Audiosignals einer Frequenztransformation (z.B. Fourier-Transformation) unterzogen.<sup>80</sup>

#### 5.1 Fourier Transformation

Die Fourier - Transformation wurde von Jean Baptiste Joseph Fourier<sup>81</sup> entwickelt und ist ein Verfahren der Signalverarbeitung. Sie zerlegt Zeitfunktionen in Frequenzkomponenten oder Elementarschwingungen (Sinus - und Kosinus – Bestandteile). Durch diesen Übergang vom Zeit- in den Frequenzbereich, wird die Gesamtheit aller Frequenzen des Signals sichtbar. Das Resultat der Transformation ist ein vieldimensionaler Vektor der mittels von PCA<sup>82</sup> in Daten und Dimensionen reduziert wird.<sup>83 84</sup>

---

<sup>80</sup> Vgl. Klassifikation on Audio-Signalen. [http://medi.uni-oldenburg.de/members/thomas/SoundClassification\\_Main\\_Color.pdf](http://medi.uni-oldenburg.de/members/thomas/SoundClassification_Main_Color.pdf) [letzter Zugriff: 12.06.2009]

<sup>81</sup> französischer Mathematiker und Physiker (\* 21. März 1768 bei Auxerre ; † 16. Mai 1830 in Paris )

<sup>82</sup> siehe 5.4, PCA=Principal Component Analysis

<sup>83</sup> Vgl. Music Information Retrieval. <http://pages.unibas.ch/LIlab/studies/IR-FS2009/Stankowski/Music%20Information%20Retrieval%20pdf.pdf>

<sup>84</sup> Vgl. Music Information Retrieval Automatische Genre-Klassifikation. <http://cvpr.uni-muenster.de/teaching/ss05/blockseminarSS05/downloads/Music-Ausarbeitung.pdf> [letzter Zugriff: 20.05.2009]

### 5.1.1 Arten der Fourier Transformation

- Diskrete Fourier-Transformation
- Fourier Reihe<sup>85</sup>
- Kontinuierliche Fourier-Transformation<sup>86</sup>

#### 5.1.1.1 Diskrete Fourier - Transformation<sup>87</sup>

Die Diskrete Fourier - Transformation ermöglicht eine Zerlegung von einer zeitdiskreten bzw. zeitbezogenen endlichen Schwingung in seine frequenzdiskreten Spektralanteile (Summe von Sinus- und Kosinusfunktionen). Das Ziel ist die Untersuchung bzw. Bestimmung des Frequenzspektrums. Weiterhin kann die Transformation ohne Verlust wieder umgekehrt werden d.h. vom Frequenz- wieder in den Zeitbereich, anhand der *Inverse Discrete Fourier Transform* (IDFT). Aufgrund der übermäßigen Redundanz der DFT wird häufig die *Schnelle Fourier - Transformation* verwendet. Die Ergebnisse sind dabei identisch.<sup>88</sup>

Einige Aufgaben der DFT<sup>89</sup>

- zur Bestimmung der in einem abgetasteten Signal hauptsächlich vorkommenden Frequenzen,
- zur Bestimmung der einzelnen Amplituden zu diesen Frequenzen
- zur Implementierung digitaler Filter mit großen Filterlängen

---

<sup>85</sup> Entwicklung einer Funktion in eine Summe von Sinus- und Kosinusfunktionen

<sup>86</sup> kontinuierliche, nichtperiodische Signale werden in ein kontinuierliches Spektrum zerlegt

<sup>87</sup> Herleitung der diskreten Fourier - Transformation – siehe Anhang

<sup>88</sup> Vgl. Music Information Retrieval Automatische Genre-Klassifikation. <http://cvpr.uni-muenster.de/teaching/ss05/blockseminarSS05/downloads/Music-Ausarbeitung.pdf> [letzter Zugriff: 20.05.2009]

<sup>89</sup> Vgl. Diskrete Fourier -Transformation [http://www.computerbase.de/lexikon/Diskrete\\_Fourier-Transformation](http://www.computerbase.de/lexikon/Diskrete_Fourier-Transformation)



### 5.1.1.2 Schnelle Fourier - Transformation (englisch: Fast Fourier Transformation)

Die „Fast-Fourier-Transformation“ wurde erstmals 1965 von James W. Cooley und John M. Tukey vorgestellt, als ein Verfahren, welches speziell für große Transformationslängen die Berechnung der DFT stark beschleunigte und weiterhin ein weiträumiges Anwendungsfeld ermöglichte. Die FFT ist eine der Standardverfahren der digitalen Signalverarbeitung.<sup>90</sup> Wie in Kapitel 3 beschrieben, existiert eine Wellenform für Audiodateien und diese besteht aus Sinusschwingungen. Die Zeitfunktion der Schwingung wird während des Verfahrens in einzelne Frequenzen zerlegt, aufgrund dessen wird die Zusammensetzung des Klanges sichtbar. Rein mathematisch führt die FFT die identische Rechnung durch wie die DFT. Der Unterschied besteht in der verminderten Zahl von Rechenoperationen. Die FFT basiert auf dem „teilerherrsche“ (engl. divide-and-conquer)<sup>91</sup> Prinzip.<sup>92</sup>

Der Algorithmus der FFT wird auf kleine Zeitausschnitte angewandt und die Ergebnisse in Blöcken nebeneinander dargestellt.

## 5.2 Wavelet Transformation

Ein weiteres Verfahren zur Signalverarbeitung ist die *Wavelet Transformation*. Hierbei wird das Frequenzspektrum zeitlich lokal betrachtet, d.h. das Signal wird mit zeitlich lokalisierten Wellen (Wavelets) gescannt, im Gegensatz zur Fourier – Transformation, die theoretisch unendliche Sinus- und Kosinusschwingungen verwendet. Die *Wavelets* sind dabei die Basisfunktionen, welche mit dem analysierten Signal verglichen werden.<sup>93 94</sup>

---

<sup>90</sup> Digitale Signalverarbeitung mit MATLAB Grundkurs mit 16 ausführlichen

<sup>91</sup> Problem der Größe  $n$  in zwei Hälften der Größe  $n/2$  Teilen

<sup>92</sup> Vgl. Eppelt, Roland M: Analyse von Audiosignalen nach musikalischen Kriterien. <http://roland.eppelt.de/da/Diplomarbeit.pdf> [letzter Zugriff: 20.06.2009]

<sup>93</sup> Vgl. Music Information Retrieval Automatische Genre-Klassifikation. <http://cvpr.uni-muenster.de/teaching/ss05/blockseminarSS05/downloads/Music-Ausarbeitung.pdf> [letzter Zugriff: 20.05.2009]

<sup>94</sup> Vgl. Eppelt, Roland M: Analyse von Audiosignalen nach musikalischen Kriterien. <http://roland.eppelt.de/da/Diplomarbeit.pdf> [letzter Zugriff: 20.06.2009]

## Arten der Wavelet Transformation

- Kontinuierliche Wavelet-Transformation
- Diskrete Wavelet-Transformation
- Schnelle Wavelet-Transformation

## 5.3 Diskrete Kosinus Transformation

Die DCT transformiert von der zeitbasierten in die frequenzbasierte Darstellung, allerdings werden nicht alle Frequenzen gleich behandelt. Das Resultat ist eine Redundanzreduktion. Hochfrequente Änderungen zwischen den Werten werden entfernt. Die berechneten Koeffizienten sind unkorreliert und besitzen eine eigene Aussagekraft. Während der Transformation löschen sich die Sinus-Anteile, sodass nur die Kosinus-Anteile übrig bleiben. Die Abbildungen 17 und 18 zeigen das Signal welches in den Raum der Cosinus-Frequenzen transformiert wird.

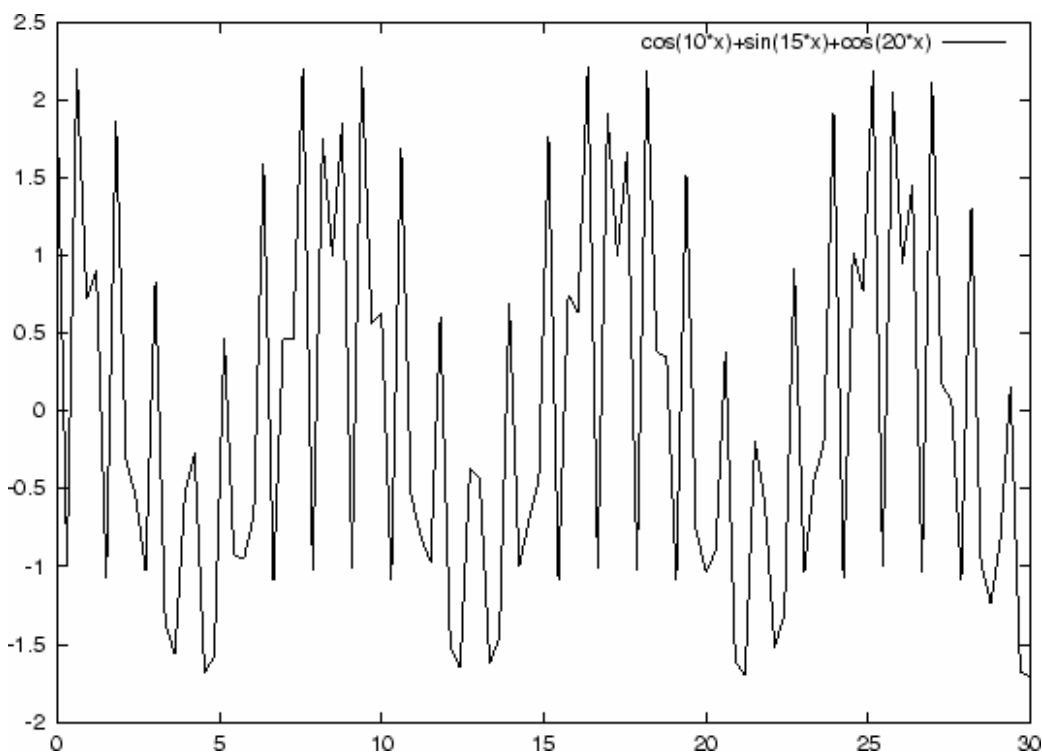


Abbildung 16 : Signal vor der Transformation

Quelle: <http://fsinfo.cs.uni-sb.de/~lynx/uni/fopra/node21.html>

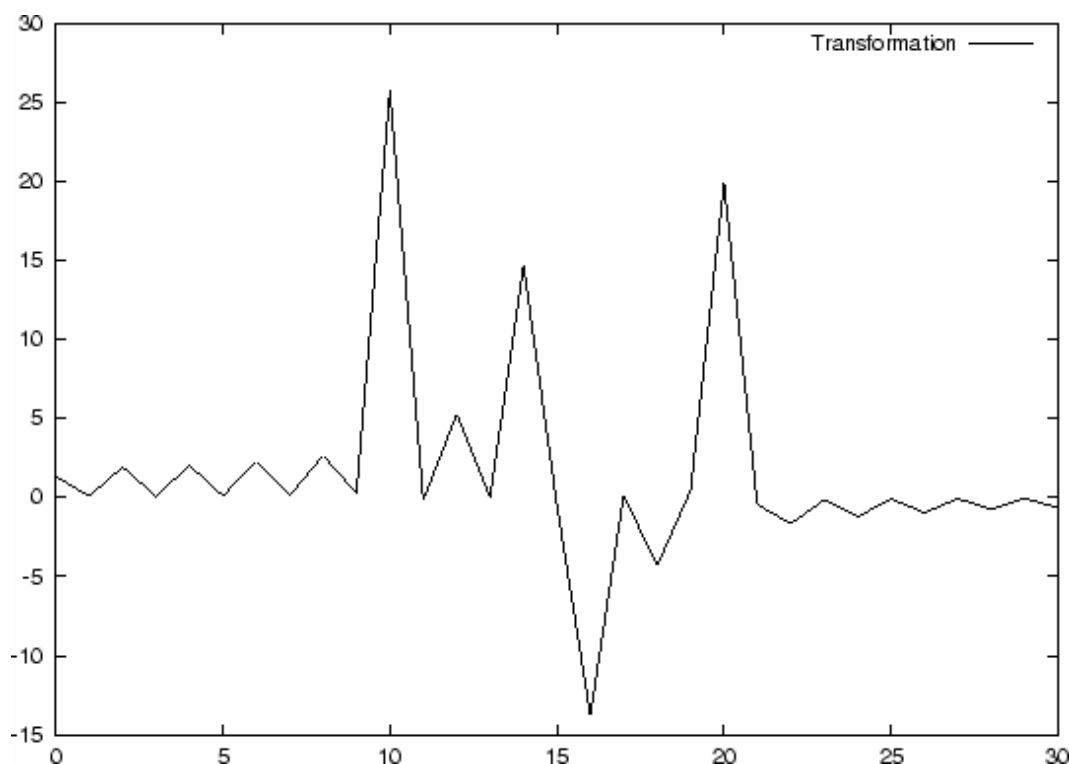


Abbildung 17 : nach der Transformation / Spitzen zeigen Cosinus Frequenzen

Quelle: <http://fsinfo.cs.uni-sb.de/~lynx/uni/fopra/node21.html>

#### 5.4 Hauptkomponentenanalyse (englisch: Principal Component analysis)

Die Hauptkomponentenanalyse wird zur Dimensionsreduktion und zur Clusteranalyse verwendet. Das Ziel ist die Reduktion von umfangreichen Datensätzen mit minimalem Informationsverlust. Durch eine angemessene Dimensionsreduktion können Klangähnlichkeiten in kompakter Form grafisch dargestellt werden.

PCA ist eine Analysemethode, um Korrelationen zwischen diversen Variablen in einem Datensatz zu bestimmen. Diese Variablen werden in unkorrelierte Variablen transformiert. Das Resultat sind die *Hauptkomponenten*. Diese stellen beispielsweise spektrale Grundcharakteristiken der Klänge dar. Somit könnte die Klangfarbe eines Instrumentes beschrieben und Klangunterschiede extrahiert werden.<sup>95</sup>

<sup>95</sup> Vgl. Siegemund, Jan: Hauptkomponentenanalyse. <http://www-mmdb.iai.uni-bonn.de/lehre/proprak0304/siegemund.pdf> [letzter Zugriff: 23.06.2009]

## 5.5 Merkmale der Merkmalsextraktion (englisch : Feature Extraction)

Die Merkmale der *Feature Extraction* werden zur Klassifikation von Audiodateien genutzt. Mit diesen „Features“ können Musikgenre eingeordnet werden, dabei unterscheidet man in *Features im Zeitbereich* und *Features im Frequenzbereich*. Erstere werden anhand des zeitlichen Verlaufs des Audiosignals berechnet, währenddessen die Features im Frequenzbereich mit Hilfe der Fourier - Transformation berechnet werden.

Spektrale Merkmal (Frequenzbereich)	Merkmale im Zeitbereich
Cepstral Coefficients	Zero Crossing Rate
Fundamental Frequency	Short Time Energy
Mel-Frequency Cepstral Coefficients	Low Energy
Spectral Centroid	
Spectral Fluctation	
Spectral Rolloff	
Spectral Bandwith	
Spectral Peaktrack	
Spectral Flatness	
Spectral Flux	
Spectral Crest Factor	
Spectral skewness	
Spectral Kurtosis	

Tabelle 4 : Überblick von Merkmalen (nicht alle Merkmale werden in dieser Arbeit beschrieben)<sup>96</sup>

### 5.5.1 Timbral Texture Features

Abgeleitet von „timbre“ dem englischen Begriff für die Klangfarbe, dienen diese Merkmale der Erfassung von Tonumfang und Brillianz des Klanges und werden rechnerisch erfasst. Alle Merkmale dieser Art basieren auf der FFT. Das Signal wird in kleine, auch möglicherweise überlappende, Segmente (auch „analysis window“ genannt) zerlegt und jedes dieser Segmente wird separat verarbeitet. Die daraus extrahierten Merkmale werden in einem

<sup>96</sup> Formeln - siehe Anhang

„Texture Window“ mit Mittelwert<sup>97</sup>, Varianz<sup>98</sup>, Mittelwert und Standardabweichung der Ableitung zusammengefasst.<sup>99</sup>

### 5.5.1.1 Spectral Centroid

Das *Spectral Centroid* wird häufig als Merkmal zur Klassifizierung unterschiedlicher Musikgenres verwendet und spiegelt Informationen über die „Fülle“ bzw. die Breite und Höhe eines Klangs wider. Man ermittelt, wo sich im Spektrum des Signals der Schwerpunkt befindet.<sup>100 101</sup>

### 5.5.1.2 Spectral Rolloff

Dieses Merkmal gibt wieder, in welcher Höhe im Spektrum ein bestimmter Energieanteil steht. Es besitzt weiterhin eine enorme Aussagekraft bei der Analyse monophoner Musiksignale. Dabei transportiert es wichtige Informationen über den Obertongehalt von Klängen. Der SR korreliert sehr stark mit dem SC, wenn das Musikstück eine hohe Komplexität an Polyphonie aufweist.<sup>102</sup>

### 5.5.1.3 Spectral Flux

Der *Spectral Flux* betrachtet nicht nur ein Spektrum, sondern mehrere aufeinander folgende spektrale Frames. Dabei gibt er Veränderungen von Block zu Block wieder.<sup>103</sup>

---

<sup>97</sup> = mittlerer Wert der Kenngröße über alle Intervalle des Signals

<sup>98</sup> = mittlere quadratische Abweichung der Kenngröße zu deren Mittelwert (wiederum über alle Intervalle des Signals)

<sup>99</sup> Vgl. Music Information Retrieval Automatische Genre-Klassifikation. <http://cvpr.uni-muenster.de/teaching/ss05/blockseminarSS05/downloads/Music-Ausarbeitung.pdf> [letzter Zugriff: 20.05.2009]

<sup>100</sup> Vgl. Projektbericht Feature Extraction Audio A. [http://www.hobbyist.de/dsp/fae\\_anzengruber.pdf](http://www.hobbyist.de/dsp/fae_anzengruber.pdf) [letzter Zugriff: 24.06.2009]

<sup>101</sup> Vgl. Robuste Genre-Klassifikation. <http://archiv.tu-chemnitz.de/pub/2008/0018/data/Klausing-Genreklassifikation.pdf> [letzter Zugriff: 22.06.2009]

<sup>102</sup> Vgl. Robuste Genre-Klassifikation. <http://archiv.tu-chemnitz.de/pub/2008/0018/data/Klausing-Genreklassifikation.pdf> [letzter Zugriff: 22.06.2009]

<sup>103</sup> ebenda

#### 5.5.1.4 Spectral Bandwith

Der *Spectral Bandwith* beschreibt die Frequenzbreite eines Signals. Berechnet wird er über eine Differenzbildung zwischen höchster und niedrigster Frequenz der Spektralkomponenten. Werte die sehr nah Null sind, werden ignoriert. <sup>104</sup>

#### 5.5.1.5 Zero crossing rate (Null-Durchgangs-Rate)

ZCR ist ein Maß für die Periodizität eines Signals. Dieses Merkmal zeigt die Anzahl der Durchgänge des Signals durch den Nullpunkt (schneiden der x-Achse), innerhalb eines Intervalls. Sollte die Rate sehr hoch sein, dominieren hohe Frequenzen das Signal und es kann davon ausgegangen werden, dass es sich um Musik handelt, da der ZCR-Wert bei Sprache niedrig angesiedelt ist. <sup>105 106</sup>

#### 5.5.1.6 Mel<sup>107</sup>-Frequency Cepstral Coefficients (deutsch: Tonheit-Frequenz-Cepstrum- Koeffizienten)

Die Einheit „Mel“ bezeichnet die Tonheit und beschreibt die wahrgenommene Tonhöhe. MFCC wird häufig in der Spracherkennung verwendet, da hauptsächlich für den Menschen wahrnehmbare Frequenzanteile verarbeitet werden. Eine Verdopplung des Mel Wertes entspricht der Verdopplung der wahrgenommenen Tonhöhe. Im Retrieval dienen die MFCC zur Beschreibung der Klangfarbe eines Musikstückes bei der Ähnlichkeitssuche<sup>108</sup>. Das Prinzip zur Erstellung von solchen Koeffizienten besteht aus mehreren Schritten<sup>109</sup>.

---

<sup>104</sup> Vgl. Projektbericht Feature Extraction Audio A. [http://www.hobbyist.de/dsp/fae\\_anzengruber.pdf](http://www.hobbyist.de/dsp/fae_anzengruber.pdf) [letzter Zugriff: 24.06.2009]

<sup>105</sup> ebenda

<sup>106</sup> Vgl. Projektbericht Feature Extraction Audio A. [http://www.hobbyist.de/dsp/fae\\_anzengruber.pdf](http://www.hobbyist.de/dsp/fae_anzengruber.pdf) [letzter Zugriff: 24.06.2009]

<sup>107</sup> Maßeinheit für die psychoakustische Größe

<sup>108</sup> Vgl. Genre-Klassifikation. <http://archiv.tu-chemnitz.de/pub/2008/0018/data/Klausing-Genreklassifikation.pdf> [letzter Zugriff: 22.06.2009]

<sup>109</sup> siehe Anhang

### 5.5.1.7 RMS-Energy (Root Mean Square = quadratischer Mittelwert)

RMS ist ein Maß für die Energie eines Signals bzw. der Lautstärke.

### 5.5.1.8 Low-Energy Feature

Das Merkmal *Low-Energy* misst in allen Analysefenstern den *RMS - Energy Wert* und berechnet diesen als Prozentsatz und zeigt die Variationen in der Lautstärke.

## 5.5.2 Rhythmic Content Features

Die Daten werden mit der Wavelet Transformation bearbeitet und auf rhythmische Regelmäßigkeiten untersucht. Das Resultat ist eine Zerlegung in 12 Subbändern, für die Erstellung des *Beat - Histogramm* zur Analyse der Rhythmik eines Liedes.<sup>110</sup>

### Beat – Histogramm

Die Darstellung des Beat Histogramms dient der Erkennung des Taktschlages. Dabei wird jeder mögliche Taktschlag-Wert erfasst. Die Spitzen entsprechen den stärkeren Taktschlägen.

### Folgende Werte werden berechnet<sup>111</sup>

- „*Beat-per-minute*“ der beiden Spitzen
- relative Amplitude der beiden Spitzen bezüglich der Summe aller Amplituden
- Verhältnis der Amplituden der beiden Spitzen zueinander
- Summe aller Werte des Histogramms

---

<sup>110</sup> Vgl. Extraction of Rhythmical Features of Audio Signals. <http://cvpr.uni-muenster.de/teaching/ws05/seminarWS05/downloads/ExtractionRhythmicalFeaturesAudioSignals-Ausarbeitung.pdf> [letzter Zugriff: 24.06.2009]

<sup>111</sup> Vgl. Genre-Klassifikation. <http://archiv.tu-chemnitz.de/pub/2008/0018/data/Klausing-Genreklassifikation.pdf> [letzter Zugriff: 22.06.2009]

## Schritte zum Beat Histogramm<sup>112</sup>

1. Wavelet-Transformation: Zerlegung des Signals in mehrere Frequenzbänder
2. Full-Wave-Rectification für einzelne Bänder
3. Low-Pass-Filtering
4. Downsampling
5. Mean-Removal
6. Aggregation
7. Enhanced-Autocorrelation

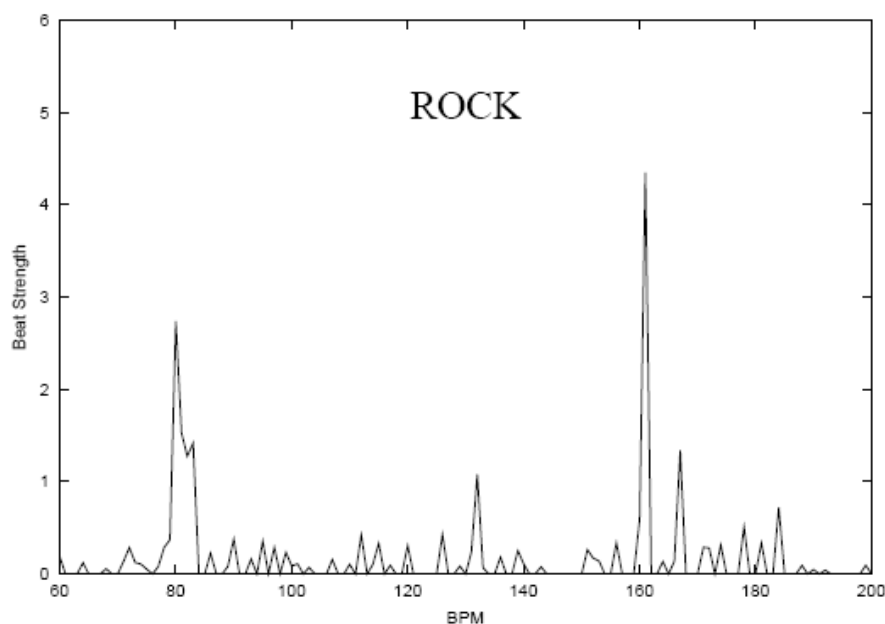


Abbildung 18 : Beispiel Beat-Histogramm (BPM=beat per minute)

Quelle: <http://webhome.cs.uvic.ca/~gtzan/work/pubs/waspaa03gtzan.pdf>

### 5.6. Chroma Features (deutsch: Chroma Merkmale)

Chroma (=Tonfarben) Merkmale codieren grob den Harmonieverlauf eines Musikstücks und beschreiben den tonalen Charakter. Das menschliche Gehör unterteilt, sich unterscheidende Halbtöne in einer Oktave, in Klassen einer „Farbigkeit“. Zum Beispiel werden  $C_0, C_1, C_2, \dots$  in der Halbtonklasse C eingeordnet. Durch diesen Effekt besteht die Möglichkeit, den Gesang eines Künstlers in einer anderen Oktave nachzusingen. Wenn jedoch die Werte der

<sup>112</sup> Vgl. Music Information Retrieval Automatische Genre-Klassifikation. <http://cvpr.uni-muenster.de/teaching/ss05/blockseminarSS05/downloads/Music-Ausarbeitung.pdf> [letzter Zugriff: 20.05.2009]



---

Halbtonklassen aufaddiert werden, erhält man die so genannten *Chroma Merkmale*. Diese werden als *Chromagramme* dargestellt.<sup>113 114 115</sup>

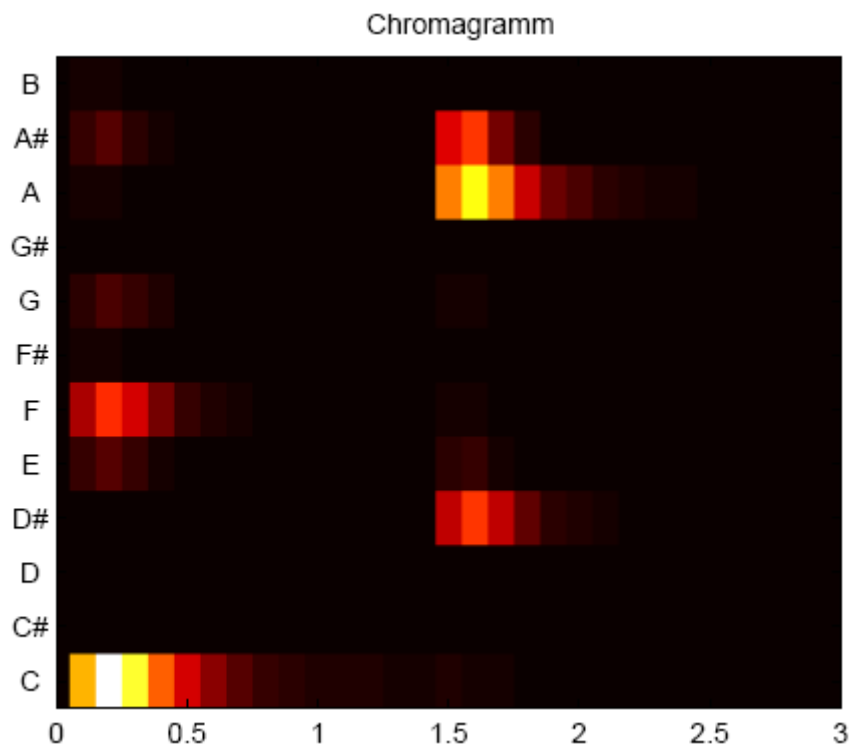


Abbildung 19 : Beispiel Chromagramm – Frequenzverlauf (x-Frequenz, y-Oktaven)

Quelle: <http://www.mpi->

[inf.mpg.de/~mmueller/publications/2008\\_Diplomarbeit\\_EwertSebastian\\_MusikSync.pdf](http://www.mpi-inf.mpg.de/~mmueller/publications/2008_Diplomarbeit_EwertSebastian_MusikSync.pdf)

---

<sup>113</sup> Vgl. Effiziente Methoden zur hochauflösenden Musiksynchronisation. [http://www.mpi-inf.mpg.de/~mmueller/publications/2008\\_Diplomarbeit\\_EwertSebastian\\_MusikSync.pdf](http://www.mpi-inf.mpg.de/~mmueller/publications/2008_Diplomarbeit_EwertSebastian_MusikSync.pdf) [letzter Zugriff: 12.06.2009]

<sup>114</sup> Beiträge zum Entwurf und Vergleich von robusten Audiomeerkmalen für das Musikretrieval. [http://www.mpi-inf.mpg.de/~mmueller/publications/2008\\_Diplomarbeit\\_KreuzerSebastian\\_AudioMatching.pdf](http://www.mpi-inf.mpg.de/~mmueller/publications/2008_Diplomarbeit_KreuzerSebastian_AudioMatching.pdf) [letzter Zugriff: 24.06.2009]

<sup>115</sup> Schritte bis zum Chromagramm – siehe Anhang

## Kapitel 6 : Mustererkennung

### 6. Verfahren zur automatischen Mustererkennung (englisch : Pattern Recognition)

Die Verfahren zur Mustererkennung ermöglichen beispielsweise eine Klassifikationen nach Musikgenre. Ihre manuelle Erstellung ist äußerst zeitaufwendig, darum werden Verfahren verwendet, um diese zu automatisieren. Der Sinn der Mustererkennung ist das Finden von Merkmalen, die eine Kategorie von einer anderen unterscheidet, beispielsweise ein Musikgenre. In der vorliegenden Arbeit nur eine bestimmte Auswahl von Lernverfahren bzw. Ansätzen geschildert. Es wird unterschieden zwischen überwachten und unüberwachten Lernverfahren (siehe Tabelle)

#### Beispiele<sup>116</sup>

Überwachte Lernverfahren	Unüberwachte Lernverfahren
Diskriminanzanalyse <ul style="list-style-type: none"> <li>• Lineare Klassifikation</li> <li>• Polynomklassifikation</li> <li>• Support Vektor Machines</li> <li>• Parameterschätzung für bekannte Dichtefunktion</li> <li>• (K) Nächster Nachbar</li> <li>• Kerndichteschätzung</li> </ul>	Clusteranalyse
	Kohonenkarte
	Art-Map

Tabelle 5: Beispiele für überwachte und unüberwachte Lernverfahren

Der Unterschied besteht darin, dass bei den unüberwachten Lernverfahren weder Klassenzuordnung noch eine Klassenzahl gegeben sind. Hingegen bei den überwachten Lernverfahren die *Soll-Klasse* bekannt ist.

<sup>116</sup> Vgl. Übersicht Lernverfahren. <http://www.uni-siegen.de/fb12/ezls/lehre/dbv2/seminar/05-kurzbeschreibunglernverfahren.pdf> [letzter Zugriff: 24.06.2009]

Die vorgestellten Klassifikationsalgorithmen haben alle die Aufgabe einen Vektor einer Klasse zuzuordnen, die bei der Merkmalsextraktion gewonnen wurden. Die einfachen Algorithmen können nur zwischen zwei Klassen unterscheiden, aber aufgrund von Erweiterungen ist auch die Unterscheidung mehrerer Klassen möglich.

## 6.1 Überwachte Lernverfahren

Bei den aufgeführten Verfahren handelt es sich um Klassifikatoren, die durch „Lernen“ auf eine bestimmte Klasse (Musikgenre) eingestellt werden. Dabei dienen die eingespeisten Trainingsdaten als Datenvektoren mit bekannten Klasseninformationen. Der Klassifikator „lernt“ sich auf die bestimmten Klassen einzustellen. Folgende Schritte müssen bei jedem Klassifikationsverfahren eingesetzt werden, um eine Klassifikation zu erstellen:

Schritte:

- Erstellung eines Trainingsdatensatz
- Analyse des Datensatzes
- Auswahl des Klassifikators
- Training des Klassifikators
- Abschätzen auf Genauigkeit
- Klassifikator ist funktionstüchtig und einsatzbereit

### 6.1.1 k-Nearest-Neighbors

Bei dem *k-Nearest-Neighbors* handelt es sich um ein instanzbasiertes Klassifikationsverfahren, bei dem eine Klassenzuordnung unter Berücksichtigung seiner *k* nächsten Nachbarn vorgenommen wird. Das Lernen besteht aus simplem Abspeichern der Trainingsbeispiele. Die Merkmalsvektoren werden als Punkte im *n*-dimensionalen Raum betrachtet. Die Abstandsfunktion ist das Maß zur Bestimmung der Ähnlichkeit zweier Punkte zueinander. Es bestimmt die Distanz zwischen Punkt A und B im Merkmalsraum. Das Verfahren ist recht simpel. Zunächst erfolgt die Abspeicherung der Beispielobjekte. Dabei werden unbekannte Objekte mittels ihrer Attributwerte klassifiziert und auf Ähnlichkeit mit den bereits gespeicherten Objekten geprüft. Das Objekt wird dann der Klasse zugeordnet, mit der es eine hohe Ähnlichkeit aufweist. Die Trainingsphase mit diesem Klassifikator verläuft recht schnell. Leider ist die Auswahl und das Speichern des Prototyps sehr speicher- und rechenintensiv<sup>117</sup>.

### 6.1.2 Support Vector machines

SVM sind ein Klassifikationsverfahren, welches auf Klassentrennung anhand einer *Hyperebene* basiert. Die Klassen werden hierbei durch eine möglichst weite Trennweite (*margin*) separiert, (siehe Abbildung 21) um später neue unbekannte Objekte möglichst korrekt zu klassifizieren. Die Weite der Ebene wird durch die am Nächsten liegenden Vektoren bestimmt, der „*Support Vectors*“ (Stützvektoren – in Abbildung markiert). Die Vektoren widerspiegeln Daten, die nach Merkmalen klassifiziert und in 2 Klassen eingeordnet wurden. Die Daten stehen als Punktvektoren im Vektorraum. SVM eignen sich sehr gut für Klassifikationsaufgaben mit wenig Trainingsmaterial, sind jedoch bei einer großen Menge von Datensätzen sehr zeitaufwendig.<sup>118</sup>

---

<sup>117</sup> Vgl. Genre-Klassifikation. <http://archiv.tu-chemnitz.de/pub/2008/0018/data/Klausing-Genreklassifikation.pdf> [letzter Zugriff: 22.06.2009]

<sup>118</sup> Vgl. FACTORS IN AUTOMATIC MUSICAL GENRE CLASSIFICATION OF AUDIO SIGNALS. <http://webhome.cs.uvic.ca/~gtzan/work/pubs/waspaa03gtzan.pdf> [letzter Zugriff: 10.06.2009]

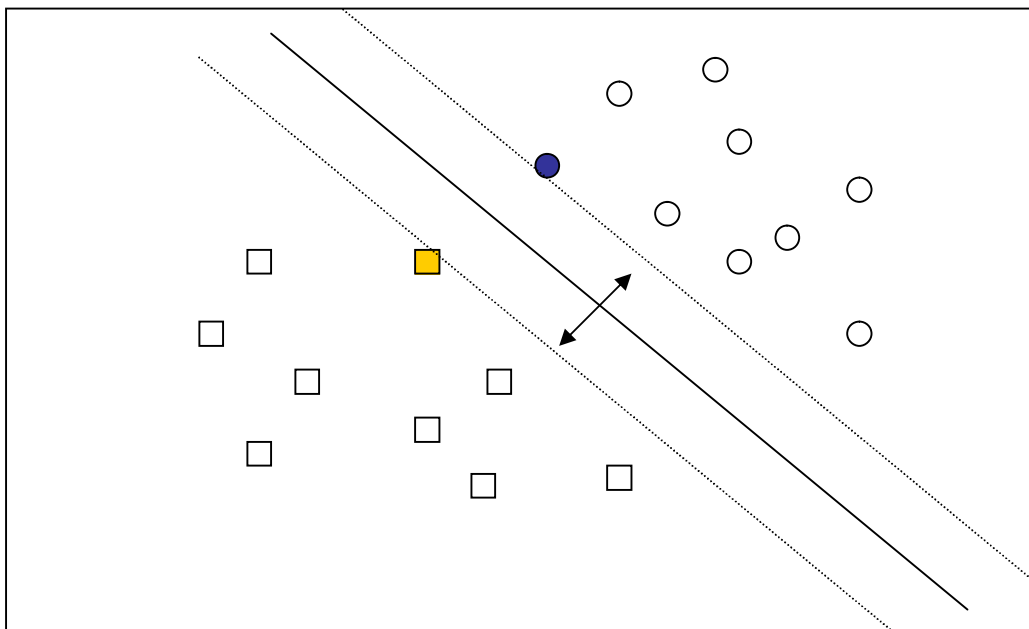


Abbildung 20 : Support Vector Machines (einfache Darstellung)

### 6.1.3 Hidden Markov Model

Das HMM ist ein stochastisches Verfahren mit 2 Zufallsprozessen, das auf Wahrscheinlichkeitsberechnungen beruht. Intern arbeiten die HMM mit einem Automaten, aufgrund dessen können zeitliche Abfolgen von Beobachtungen in die Klassifikation miteinbezogen werden. Musik als auch Sprache besitzen zeitliche Abhängigkeiten. Somit sind diese keine zufällige Aneinanderreihung von Tönen und können identifiziert werden. Der erste Zufallsprozess bzw. der Kern dieses stochastischen Prozesses (=eine Folge von elementaren Zufallsereignissen) ist die so genannte versteckte („hidden“) „Markov Kette“. Dieser Prozess zeigt, dass alle zukünftigen Zustände von dem momentan herrschenden Zustand abhängen. Dabei besteht der Prozess aus mehreren Schritten. Jeder dieser Schritte wird durch einen Zustandswechsel beschrieben. Sollte dabei ein Zustand erreicht werden, wird ein Ausgabesymbol mit einer bestimmten Wahrscheinlichkeit erzeugt, dies entspricht dem zweiten Zufallsprozess. Jeder Zustand wird bei der Ausgabe zu geordnet, abhängig vom aktuellen Zustand. Die Ausgabe ist somit eine Sequenz von Ereignissen. HMM wird in der Spracherkennung verwendet und für die Suche von Phonemen genutzt, sowie Musikklassifikationen.

Dabei kommen als Beobachtungen die Merkmalsvektoren zum Tragen. Jedes Genre, ob nun Pop oder Jazz, besitzt ein eigenes Hidden Markov Modell. Dabei wird der *Baum-Welch-Algorithmus* eingesetzt. Dieser ist in der Lage auf Übergang- und Beobachtungswahrscheinlichkeiten zuzuschließen, wobei dafür die Anzahl der Zustände und Beobachtungen eine Rolle spielt. Auf diese Weise werden die Modelle der Genre trainiert.

#### 6.1.4 Gaussian Mixture Models

Bei dem GMM wird davon ausgegangen, dass die zeitliche Information in den Merkmalvektoren keine bzw. eine untergeordnete Rolle spielt.

Die Modelle sind Mischungen multivariater Normalverteilungen, die durch Mittelwerte, Kovarianzmatrizen und Gewichtungen parametrisiert werden. Die Parameter werden dann mit einem speziellen Algorithmus geschätzt, dem *Expectation Maximization-Algorithmus*. Das Resultat ist ein Modell, welches eine maximale Wahrscheinlichkeit für Trainingdaten bietet. Dabei wird für jedes Genre ein eigenes Modell trainiert (ähnlich HMM), die zu einem Klassifikator zusammengesetzt werden. Diese Methode ist nicht sehr zeitaufwendig, deswegen vergleichsweise populär.<sup>119</sup>

### 6.2 Transformationen im Merkmalsraum

Das Ziel von Merkmalsraumtransformationen ist die Reduzierung von Dimensionalität der Merkmalsvektoren. Damit wird die Berechnungsdauer verkürzt.

#### 6.2.1 Hauptkomponentenanalyse

Die Hauptkomponentenanalyse spiegelt, dreht und skaliert die Achsen eines Systems in die Richtung seiner minimalen Varianzen, welches eine bessere

---

<sup>119</sup> Vgl. Expectation Maximization Algorithmus und Modellschätzung. <http://www-mmdb.iai.uni-bonn.de/lehre/proprak0304/mueller.pdf> [letzter Zugriff: 01.06.2009]

Separierung der Klassen zur Folge hat. Den Ausgangspunkt bilden dabei die aus der Merkmalsextraktion gewonnenen Merkmalsvektoren.<sup>120</sup>

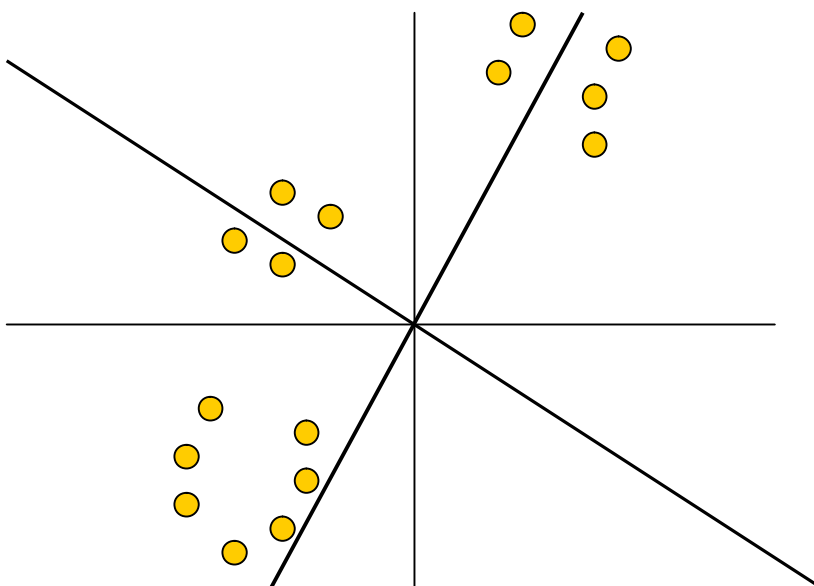


Abbildung 21 : Principal Component analysis

### 6.2.2 Lineare Diskriminanzanalyse (englisch: linear discriminant analysis)

Die lineare Diskriminanzanalyse berücksichtigt in besonderem Maße die Klassenzugehörigkeit der einzelnen Daten. Folgende Bedingungen müssen bei einer LDA erfüllt werden.

1. kompakte Darstellung der Klassen im transformierten Raum
2. Trennung der Klassen durch ausgewählte Richtungen im transformierten Raum

Die Ziele der LDA sind denen der PCA sehr ähnlich. Beide dienen der Dimensionsreduzierung und als Klassifikator (Diskriminanzfunktion) zur Darstellung von Merkmalen. Eine Diskriminanzfunktion ermöglicht eine Trennung von Klassen und überprüft die Bedeutung der Merkmalsvariablen (siehe SVM).<sup>121</sup>

---

<sup>120</sup> Vgl. PCA. <http://www.digimusik.de/PCA/bsp3d.html> [letzter Zugriff: 20.06.2009]

<sup>121</sup> Vgl. Lineare Diskriminanzanalyse. [http://www.statistics4u.info/fundstat\\_germ/cc\\_lda\\_intro.html](http://www.statistics4u.info/fundstat_germ/cc_lda_intro.html) [letzter Zugriff: 10.06.2009]

## Kapitel 7 : Musiksynchronisation

### 7.1 Bedeutung von Musiksynchronisation für Musikbibliotheken

Arten von Musikbibliotheken<sup>122</sup>

#### Öffentliche Musikbibliotheken

- breites Angebot an Musikalien, Musikbüchern und Musikzeitschriften, Tonträgern und Musikvideos aus allen Bereichen der Musik
- größere Bibliotheken besitzen zusätzlich wissenschaftliche Editionen und entsprechende Sekundärliteratur, sowie Nachschlagewerke

#### Wissenschaftliche Musikbibliotheken

- sind Musikabteilungen an Wissenschaftlichen Bibliotheken
- besitzen Musikliteratur und Musikalien
- Quellenmaterialien: Musikhandschriften, alte Drucke, Nachlässe, Briefe
- Sammlungen mit audiovisuellen Medien, hauptsächlich Tonträger

#### Musikhochschulbibliotheken

- Arbeiten selbstständig
- Wissenschaftliche Orientierung
- Betreuung von Aufführungsmaterialien
- Nutzung ausschließlich für Hochschulangehörige

#### Rundfunk- und Orchesterbibliotheken

- nur für Mitarbeitern der Rundfunkanstalten bzw. Mitgliedern der Sinfonie- und Opernorchester verfügbar
- Einteilung in die Bereiche: (Musik-)Bibliothek, Notenarchiv und Schallarchiv

---

<sup>122</sup> Musikinformation und Musikdokumentation.  
[http://www.miz.org/static\\_de/themenportale/einfuehrungstexte\\_pdf/08\\_MedienRecherche/jaenecke.pdf](http://www.miz.org/static_de/themenportale/einfuehrungstexte_pdf/08_MedienRecherche/jaenecke.pdf)  
[letzter Zugriff: 24.06.2009]



Die Bedeutung erschließt sich anhand eines Blickes auf moderne digitale Musikbibliotheken. Dort befinden sich Dokumente in zahlreichen Varianten z.B. Audiodaten oder Partiturdaten, MIDI Daten oder Gesangstexte. Die Hauptaufgabe der Synchronisation ist die automatische Verlinkung von Musikdatenströmen desselben Musikstückes in unterschiedlichen Formaten. Die Schwierigkeit besteht bei der Komplexität und Vielfältigkeit von Musikdaten.

## 7.2 Audio-Partitur Synchronisation (audio-to-score alignment)

Bei einer Audio-Partitur Synchronisation wird ein Musikstück verlinkt, welches als Audioaufnahme und als Partitur existiert. Das Ziel ist dabei das Erkennen einer bestimmten Stelle der Partitur mit der exakt selben Stelle in der Audiodatei.<sup>123</sup>



Abbildung 22 : Verlinkung von Musikdaten in verschiedenen Formaten

Quelle: <http://subs.emis.de/LNI/Proceedings/Proceedings68/GI-Proceedings.68-11.pdf>

<sup>123</sup> Vgl. Aktuelle Aspekte des Music Information Retrieval.  
Online verfügbar unter <http://www.datenbank-spektrum.de/pdf/dbs-18-24.pdf> [Letzter Zugriff: 11.06.2009]

### 7.2.1 Probleme

Die Struktur einer Audiodatei besitzt keine Ähnlichkeit mit der einer Partitur (notenbasierte Parameter). Tempo, Dynamik oder die Ausführung von Notengruppen haben somit einen weiten Interpretationsraum. Ein Musikstück auf einer CD codiert besitzt alle Parameter, die für die Wellenform notwendig sind, leider aber nicht die Notenparameter. In der Abbildung ist zusätzlich die MIDI Datei abgebildet. Sie dient als eine Art Bindeglied, denn sie enthält Partiturnformationen, als auch die Interpretationen des Titels.

Die Synchronisation ist bei komplexen Musikstücken einfacher als bei oberflächlichen. Aus einer Symphonie von Beethoven lassen sich mehr Merkmale extrahieren, als aus einem einfachen Popsong. Leider ist bisher noch keine hochauflösende Synchronisation möglich d.h. auf Taktebene lässt sich die Synchronisation durchführen, aber nicht auf Notenebene.

### 7.3 Dynamic Time warping

Unter dem *Dynamic Time Warping* versteht man eine Technik die dazu dient, 2 Zeitreihen aneinander auszurichten. Die Zeitreihen werden aus 2 Datenströmen erzeugt die jeweils eine lokal gestreckte oder gestauchte Version des Anderen sind. Dabei ist von Belang welche Elemente der Zeitreihen einander ansprechen.<sup>124 125</sup>

---

<sup>124</sup> Chon, SongHui (2007): A Study of Synchronization of Audio Data with Symbolic Data. Online verfügbar unter <http://ccrma.stanford.edu/~shchon/projects/Music254/project.pdf> [letzter Zugriff: 24.06.2009]

<sup>125</sup> Vgl. Effiziente Methoden zur hochauflösenden Musiksynchronisation. [http://www.mpi-inf.mpg.de/~mmueller/publications/2008\\_Diplomarbeit\\_EwertSebastian\\_MusikSync.pdf](http://www.mpi-inf.mpg.de/~mmueller/publications/2008_Diplomarbeit_EwertSebastian_MusikSync.pdf) [letzter Zugriff: 12.06.2009]

---

## 7.4 SyncPlayer System<sup>126</sup>

Der SyncPlayer ist eine Software zur synchronen Visualisierung der verlinkten Musikdaten, während der akustischen Wiedergabe des Audiosignals.

### Funktionen

- Player
- Score Following
- Audio structure
- Karaoke
- Lyrics Search
- Audio Search

Dies gelingt durch ein Visualisierungsmodul, welches in diversen Darstellungsmodi arbeiten kann.

Der SyncPlayer spielt weiterhin WAVE und MP3 ab, wobei bei einem klassischen Stück die Partitur verfolgt werden kann. Klickt man bei der Partitur auf eine bestimmte Stelle wird diese abgespielt. Die Partituren sind dabei auf einem Server abgespeichert.

---

<sup>126</sup> Vgl. SyncPlayer Homepage. <http://audentify.iai.uni-bonn.de/synchome/> [letzter Zugriff: 20.06.2009]

## Kapitel 8 : Suchmethoden

### 8.1 Überblick der Hauptsuchmethoden des Audio Retrieval<sup>127</sup>

Das Audio Retrieval setzt sich aus 4 Hauptsuchmethoden zusammen

1. Suche auf Basis von Metadaten
2. Suche auf akustischen Merkmalen
3. Suche in gesprochenem Text
4. Suche mit Noten

#### 8.1.1 Suche auf Basis von Metadaten

Die Suche auf Basis von Metadaten ist keine direkte Audio Retrieval Suchmethode, denn sie lehnt sich ausschließlich an das herkömmliche Textretrieval an. Audiodaten werden bei dieser Art der Suche nur indirekt berücksichtigt.

Gängige Eingabemöglichkeiten bzw. Attribute

- Künstler
- Album
- Musiktitel
- Genre
- ...

Das Ergebnis ist eine Trefferliste mit den gewünschten Informationen. Beispielsweise basieren file-sharing<sup>128</sup> Programme auf dieser Suchmethode, wobei auch jede gängige Suchmaschine eine Metadatenuche anbietet.<sup>129</sup>

---

<sup>127</sup> Audio Retrieval. [http://www.lemme.at/files/GWA\\_8\\_Audioretrieval.pdf](http://www.lemme.at/files/GWA_8_Audioretrieval.pdf) [letzter Zugriff: 23.06.2009]

<sup>128</sup> Austausch bzw. Weitergeben von Dateien durch Nutzer

<sup>129</sup> Audio Retrieval. [http://www.lemme.at/files/GWA\\_8\\_Audioretrieval.pdf](http://www.lemme.at/files/GWA_8_Audioretrieval.pdf) [letzter Zugriff: 23.06.2009]

### 8.1.2 Suche mit akustischen Merkmalen wie Lautstärke, Tonhöhe usw.

Bei dieser Suchmöglichkeit werden die signifikanten Signalinformationen aus der Audiodatei extrahiert und gespeichert. Das sind beispielsweise die Intensität, Lautstärke, Tonhöhe und Helligkeit. Zu diesem Zweck wird das Audiosignal in Intervalle unterteilt und zu jedem dieser Intervalle ein bestimmtes Merkmal gespeichert. Durch diesen Prozess wird ein mehrdimensionaler Vektor erhalten, mit der Repräsentation der akustischen Merkmale bzw. Eigenschaften der Melodie. Die FFT stellt die Werte in zeitabhängigen Vektorfunktionen dar, auf denen die Suche angewendet wird.<sup>130 131</sup>

### 8.1.3 Suche in gesprochenem Text

Diese Art der Suche greift auf gesprochenen Text im Audiostrom zurück. Diese Suchmöglichkeit ist nur mittels robuster Spracherkennung möglich. Die Probleme bestehen in der Erkennung von Phonemen und Wörtern. Die Anfrage wird getippt (oder gesprochen) und umgewandelt. Dieser Aspekt der Hauptsuchmethoden wird im Verlauf der Arbeit nicht weiter ausgeführt.<sup>132</sup>

### 8.1.4 Suche basierend auf Noten bzw. Tonintervallen

Diese Suchmethode betrachtet die Notenrepräsentation des Liedes, zusätzlich auch nach Takt und Tempo. Unter diese Suchmethode fällt beispielsweise das *Query by Humming* System.<sup>133</sup>

---

<sup>130</sup> Vgl. Audio Retrieval. [http://www.lemme.at/files/GWA\\_8\\_Audioretrieval.pdf](http://www.lemme.at/files/GWA_8_Audioretrieval.pdf) [letzter Zugriff: 23.06.2009]

<sup>131</sup> Vgl. Melodie-Suche n einem Archiv von historischen Notenhandschriften. [http://www.enotehistory.de/db/SA\\_KatrinZiehe\\_2006.pdf](http://www.enotehistory.de/db/SA_KatrinZiehe_2006.pdf) [letzter Zugriff: 19.06.2009]

<sup>132</sup> ebenda

<sup>133</sup> ebenda

## 8.2 Einzelne Suchmethoden

### Musiktechnologien

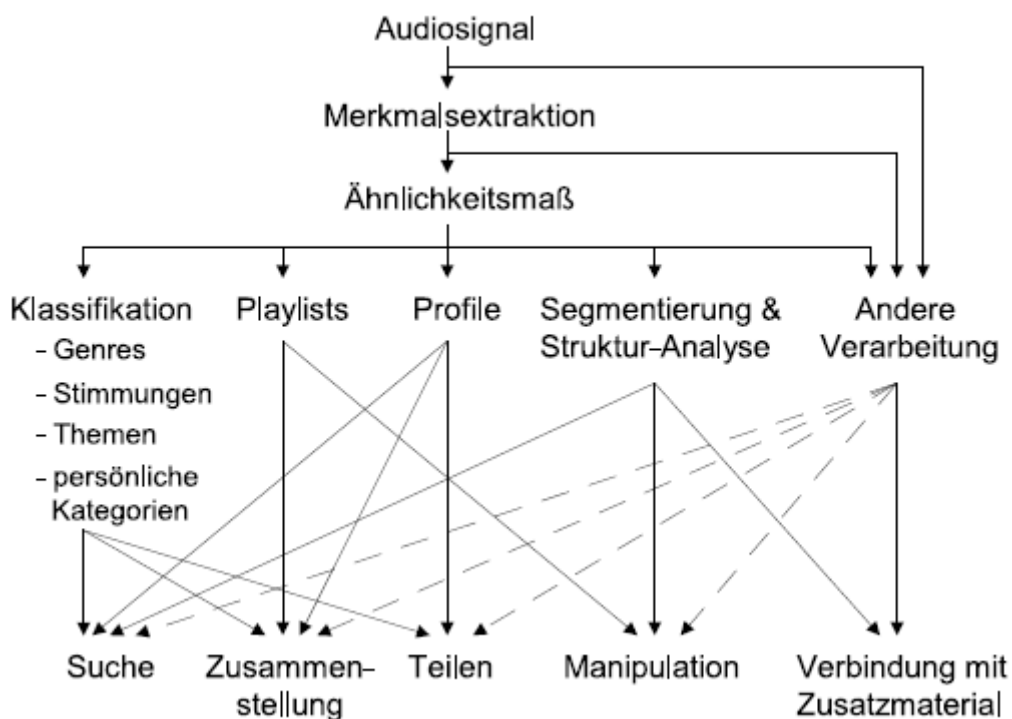


Abbildung: Technologien der Musikanalyse

Quelle: <http://www.jeroenbreebaart.com/papers/daga/daga2008.pdf>

Musiktechnologien können 5 Funktionen<sup>134</sup> unterstützen

- Suche
- Zusammenstellung
- Teilen z.B. Musikgeschmack
- Manipulation z.B. Remix
- Verbindung mit Zusatzmaterial z.B. Notation

Musik wird anhand eines Abstandsmaßes klassifiziert, welches auf zuvor extrahierten Merkmalen berechnet wird. Weitere Anwendungen sind die Erstellung von Playlists und Benutzerprofilen.

<sup>134</sup> Anwendungsbeispiele für Musikanalyse-Algorithmen.  
<http://www.jeroenbreebaart.com/papers/daga/daga2008.pdf>

### 8.2.1 Query by example

Das „Query by Example Paradigm“ wird zur Untersuchung von Ähnlichkeit genutzt. Die Anfrage besteht aus einem Musikfragment und das Ergebnis ist eine Liste, bei der das Ranking auf der inhaltlichen Ähnlichkeit basiert. Es gibt zwei Hauptansätze für Query by Example. Der erste Ansatz bezieht sich auf das Singen einer Melodie als Anfrage durch den Nutzer – Query by Humming, siehe 8.2.6. Der zweite Ansatz welcher als „Query by Example“ bezeichnet wird, nutzt als Anfrage eine Audiodatei. Das Ranking besteht aus Audiodateien anhand der inhaltlichen Ähnlichkeit. Dieser Ansatz äußert sich anhand verschiedener Algorithmen und Tools die auf der Extraktion von inhaltlichen Informationen von Audiosignalen basieren.<sup>135</sup>

### 8.2.2 Suche anhand der melodischen Kontur

Der Parsons Code repräsentiert die relative Höhenänderung von Noten einer Melodie. Die Anfragemelodie wird mit allen in der Datenbank vorliegenden Melodiecodes verglichen und gibt dem Nutzer die daraus entstehenden Resultate. Falls das passende Stück mit dem Melodiefragment nicht auffindbar ist, muss das so genannte „Substring-Matching“ durchgeführt werden. Das bedeutet, dass sämtliche Melodien gefunden werden, die das Fragment aufweisen, welches als Suchbedingung eingegeben wurde. Der Parsons Code wird von musipedia.org genutzt.

Es ist sinnvoll, die ungefähre Tonhöhenfolge der ersten 10 bis 15 Noten des gesuchten Liedes zu kennen, so wird die Suche präziser. Die Note am Anfang ist dabei der neutrale Bezugspunkt der mit einem Stern gekennzeichnet wird (siehe Tabelle).

---

<sup>135</sup> Vgl. Beyond the Query-By-Example Paradigm: New Query Interfaces for Music Information Retrieval. <http://www.cs.berkeley.edu/~andreya/papers/icmc02.pdf> [letzter Zugriff: 25.05.2009]

Einige Beispiele zur Eingabe

Lied	Beispielcode
"Twinkle Twinkle Little Star"	*rururddrdrdrd
"Silent Night"	*udduuddurdurdurudddudduruddduddurudduuddduddd

Tabelle 6 : Beispiele für den Parsons Code

Quelle: [http://www.reachinformation.com/define/Parsons\\_code.aspx](http://www.reachinformation.com/define/Parsons_code.aspx)

## 8.2.3 Nutzung von Klassifikation

Verschiedene Portale bieten eine Suche nach Genre an. Dabei unterscheidet sich die Klassifikation zwischen manueller und automatischer Klassifikation.

### 8.2.3.1 Einfache Klassifikation

Klassifikationen dienen der besseren Strukturierung von Daten und einer schnelleren Übersicht. Auch im musikalischen Bereich wird diese Möglichkeit genutzt um den Nutzern die Suche zu vereinfachen.

Viele Musikportale verwenden Klassifikationen, um den Nutzer schnell Informationen über favorisierte Musik zuzuliefern und eventuell zu empfehlen. Dabei werden die Musiktitel nach Genre geordnet, um die Navigation zu erleichtern. Musik Genre sind kategorisierte Beschreibungen zur Strukturierung der ansteigenden Mengen von verfügbarer Musik im Web. Leider können unter Genre keine scharfen Grenzen gezogen werden, da Musik immer subjektiv empfunden wird. Um dieses Problem zu lösen, ist bei einfachen Klassifikationen die auf einer Website sichtbar sind, beispielsweise eine Verdichtung möglich, wie Rock-Pop. Der Nutzer ist aufgrund dessen weniger eingeschränkt und kann seinen Musikgeschmack eventuell erweitern. Die Musiksuchmaschine „Musuma“ teilt seine gespeicherten Interpreten in Genre ein.



...
Oldies (16)
Oper (59)
Pop (47)
Progressive (38)
Psychedelic Rock (0)
Punk (251)
R and B (156)
Rap (12)
Reggae (49)
Rock (634)
Rock 'n' Roll & Rockabilly (40)
...

Abbildung 23 : Auszug der Musikgenre von Musuma

Quelle: <http://www.musuma.de/Genres/index.html>

Bei einer Suchanfrage in einer Musikdatenbank, werden die jeweiligen Genre mit angezeigt. Viele davon wurden manuell zu geordnet was einen sehr hohen Zeitaufwand darstellt. Aus diesem Grund wird auf die automatische Klassifikation zurückgegriffen.

### 8.2.3.2 Automatische Genre Klassifikation

Musikgenre besitzen charakteristische Komponenten wie z.B. die Instrumentation oder die rhythmische Struktur. Bei einer automatischen Genre Klassifikation greift ein Algorithmus zur Identifizierung. Wie in Kapitel 6 beschrieben werden zur Erstellung solcher Klassifikationen verschiedene Verfahren verwendet.<sup>136 137</sup>

---

<sup>136</sup> Vgl. Automatic Musical Genre Classification Of Audio Signals.  
<http://ismir2001.ismir.net/pdf/tzanetakis.pdf> [Letzter Zugriff: 11.06.2009]

<sup>137</sup> Vgl. Anwendungsbeispiele für Musikanalyse-Algorithmen.  
<http://www.jeroenbreebaart.com/papers/daga/daga2008.pdf> [letzter Zugriff: 11.06.2009]

### 8.2.3.3 Klassifikation anhand der Stimmung des Nutzers (englisch: mood classification)

Die „*mood classification*“ richtet sich nach dem momentanen Gemütszustand des Nutzers. Sie verbindet Musik mit den Emotionen. Ähnlich der Genre Klassifikation ist diese Klassifikation rein subjektiv, darum wird bei der Erstellung die Subjektivität so weit wie möglich eingeschränkt und die Stimmung, die ein Lied ausdrückt bestimmt. Bei einer Modellierung der Klassen werden diese berücksichtigt, die von Probanden übereinstimmend Verwendung finden. Die Auswahl von beispielsweise „fröhlich“ oder „traurig“ ist für die späteren Tests sehr wichtig, denn es werden ausschließlich Musikstücke gewählt die für einen Stimmung eindeutig identifizierbar erscheinen.<sup>138 139</sup>

### 8.2.4 Query by Tapping

Bei einem „Query by Tapping“ System wird der Rhythmus eines Liedes mit einer Taste der Tastatur „geklopft“. Es wird die rhythmische Kontur anhand der Notenlängen repräsentiert. Die Daten werden dann mit denen in der Datenbank auf Übereinstimmung oder die Wahrscheinlichkeit der Übereinstimmung verglichen. Der Algorithmus zu diesem System wurde von Studenten der Simon Fraser Universität in Vancouver entwickelt. „Songtapper“ kann bisher 80627 (aktueller Stand) Songs erkennen<sup>140</sup>.

### 8.2.5 Query by melody

Der Nutzer lässt bei dieser Methode eine MIDI (bei anderen Formaten wird transkribiert) Datei einlesen. Dann wird die Melodie mit denen der Datenbank verglichen und mit den dazugehörigen Metadaten ausgegeben. QbM ist

---

<sup>138</sup> Vgl. Content-based Mood Classification for Photos and Music.

<sup>139</sup> Vgl. Anwendungsbeispiele für Musikanalyse-Algorithmen.

<http://www.jeroenbreebaart.com/papers/daga/daga2008.pdf> [letzter Zugriff: 11.06.2009]

<sup>140</sup> Vgl. Song Search and Retrieval by Tapping. <http://www.aaai.org/Papers/AAAI/2005/ISD05-014.pdf> [letzter Zugriff: 23.06.2009]

ausschließlich möglich, wenn symbolische Daten vorhanden sind, was bei aufgenommener Musik nur sehr selten der Fall ist.<sup>141</sup>

## 8.2.6 Query by Humming – Query by Singing

*Query by Humming* ist ein Erkennungssystem für Melodien. Besonders nützlich ist dieses System bei Nutzern, die einen bestimmten Takt bzw. ein Bruchstück eines Liedes im Ohr haben, aber es nicht identifizieren können. Dabei ist das Ziel eines solchen Systems, das Wiederfinden einer Melodie, welche als Anfrage vom Nutzer in ein Mikrofon gesummt oder gesungen wird. Danach erfolgt ein Abgleich mit den gespeicherten Melodien der Datenbank. Das System zerlegt dabei die vorgesungene Melodie in einzelne Töne, die nach der Tonhöhe analysiert und quantisiert werden.

### 8.2.6.1 Anforderungen an ein Query by Humming System

Ein effizientes QBH System zeichnet sich durch diverse Faktoren aus, um dem Nutzer das bestmögliche Ergebnis zu liefern.<sup>142</sup>

- Genauigkeit des Ergebnisses
- schnelle Antwort
- umfangreiche Datenbank
- Skalierbarkeit
- Flexibilität
- Robustheit
- Brauchbarkeit/ Nutzbarkeit

---

<sup>141</sup> A Mid-level Melody-based Representation for Calculating Audio Similarity. <http://lgm.fri.uni-lj.si/~matic/clanki/ismir2006.clanek.pdf> [letzter Zugriff: 24.06.2009]

<sup>142</sup> Vgl. Query-by-humming im Internet. [http://pi3.informatik.uni-mannheim.de/downloads/hauptstudium/seminare/papers/Ausarbeitung\\_QueryByHumming.pdf](http://pi3.informatik.uni-mannheim.de/downloads/hauptstudium/seminare/papers/Ausarbeitung_QueryByHumming.pdf) [letzter Zugriff: 24.06.2009]

### 8.2.6.2 Transkription

Es gibt 2 verschiedene Arten von Transkription, die monophone und die polyphone Transkription. QbH Systeme können bisher nur Erstere verwenden. Seit den letzten 10 Jahren wurde bereits nach einem Patentrezept für die polyphone Transkription gesucht, was sich aber als äußerst schwierig herausstellte. Die polyphone Musik weist eine wesentlich höhere Komplexität auf und obwohl bereits erste Erfolge erzielt wurden, ist in unmittelbarer Zukunft noch keine Lösung des Problems in Sicht. Bei einem QbH System wird die gesummte Melodie zunächst in eine symbolische Darstellung transkribiert. Das ist ein notwendiger Schritt um eine Melodiebeschreibung zu erhalten, die in dem Format in der Datenbank vorliegt. Dabei wird das gesummte Signal zunächst einer Tonhöhenenerkennung zugeführt und anschließend einer Rhythmuserkennung. Nach diesen Prozessen werden die extrahierten Parameter ausgewertet und in eine symbolische Darstellung gebracht.<sup>143</sup>

### 8.2.7 Similarity Search (Ähnlichkeitssuche)

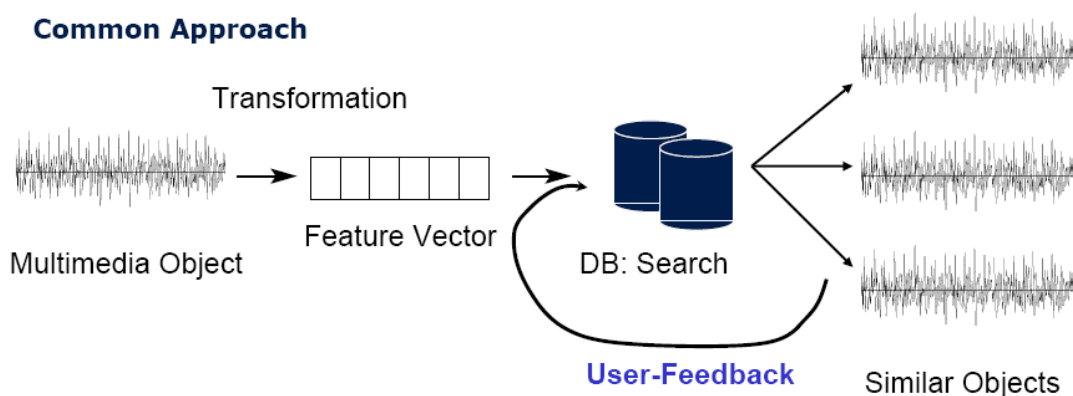


Abbildung 24 : Prozesse einer Similarity Search

Quelle: [http://www.db.inf.tudresden.de/files/research/eyes4ears/Eyes4Ears\\_MusicNetwork2005.pdf](http://www.db.inf.tudresden.de/files/research/eyes4ears/Eyes4Ears_MusicNetwork2005.pdf)

<sup>143</sup> Vgl. Batke, Johann M.: Untersuchung von Melodiesuchsystemen sowie von Verfahren zu ihrer Funktionsprüfung

In der Abbildung sind die Prozesse der *Similarity Search* dargestellt. Ein Signal wird eingegeben und in der Datenbank durchsucht, um die Ähnlichkeit zu den bereits gespeicherten Objekten zu berechnen. Anschließend wird absteigend nach dem Ähnlichkeitsmaß sortiert und eine Liste mit ähnlichen Objekten generiert.

## **Kapitel 9: Anwendungen**

### **Anwendungen von Music Information Retrieval im Internet**

Es gibt verschiedene Anwendungen im Internet, die auf MIR basieren. Insbesondere auf bestimmten Algorithmen. Diese widerspiegeln sich in Musikempfehlungssystemen und Musiksuchmaschinen. Allerdings haben viele nur eine geringe Lebensdauer, da sie nicht finanziert werden können. In der vorliegenden Arbeit wird nur auf wenige Angebote eingegangen um die Funktionalität von MIR Systemen zu demonstrieren.

Zugangsart: Alle Portale sind über die jeweilige Homepage erreichbar.

### **9.1 Musiksuchmaschinen in der Praxis**

#### **9.1.1 Suche mittels Metadaten**

Die Suche nach Metadaten wird häufig von „Peer-to-Peer Filesharing“-Programmen (z.B. Napster) verwendet und ist eine Variante des Textretrieval. Die Suchanfrage orientiert sich anhand der in der Datenbank abgelegten Attribute der jeweiligen Audiodatei.

Die Ausgabe ist eine Trefferliste, der in Frage kommenden Objekte.

Man könnte diesbezüglich jede Suchmaschine auflisten, da in der Regel jedes Portal diese Suche anbietet.

#### **9.1.2 Query by Tapping und Suche anhand der melodischen Kontur**

##### **9.1.2.1 Musipedia**

Die Musiksuchmaschine Musipedia identifiziert Musik anhand von 5 Suchmethoden.

## Klaviatur Suche

Die Eingabe erfolgt über ein virtuelles Klavier. Dabei werden die dabei gespielten Noten in einen Code umgesetzt.

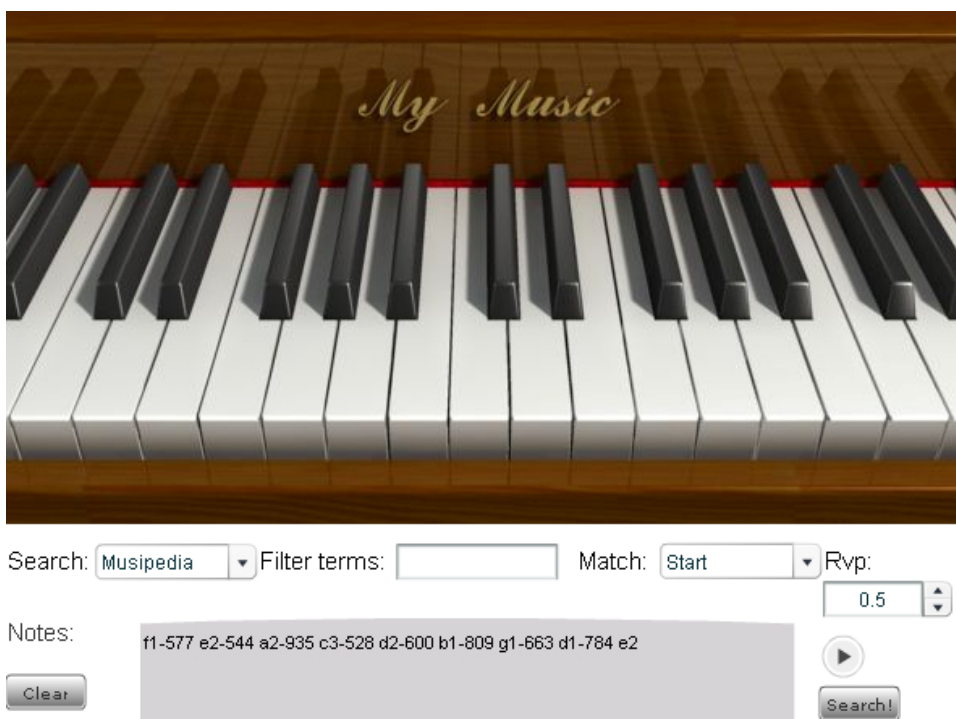


Abbildung 25 : Klaviatur Suche von Musipedia

Die Suche kann später noch verfeinert werden, indem die Tonhöhe oder der Rhythmus besser ausbalanciert wird. Bei dieser Suchmethode sollte man auf dem Gebiet Musiklehre und Musiktheorie sehr bewandert sein, ansonsten wird kein korrektes Ergebnis folgen.

## Mikrofon Suche

Die Suche ist nur mit pfeifen der Melodie durchführbar. Singen führt zu falschen Ergebnissen.

## Rhythmus Suche

Die Rhythmus Suche funktioniert anhand des „Query by Tapping“ Prinzips. Musipedia versucht für den Titel ähnliche Objekte in der Datenbank zu finden. Als Test wurde der Song „Take me out“ von Franz Ferdinand gewählt, aber leider konnte anhand des „Tappens“ kein einziges Mal der Song gefunden werden (obwohl er in der Datenbank vorhanden ist).

## Suche nach melodischer Kontur

Die Methode beruht auf dem Parsons Code.

**Melodic Contour**  
(Parsons Code):  U=Up  
D=Down  
R=Repeat

Keywords and/or  
"search phrases":

**Restrict the search to these categories:**

- Classical
- Popular
- Folk Songs
- Hymns & Carols
- National Anthems

Abbildung 26 : Eingabe des Parsons Code

Es ist von essentieller Bedeutung, den Code so lang wie möglich zu gestalten. Bei so einer kurzen Folge wie in der Abbildung wird das Ergebnis mit einer hohen Wahrscheinlichkeit ungenau sein.

## Einzeichnen von Noten

Bei dieser Methode müssen mindestens 5 Noten mit der Maus für die Suche eingegeben werden. Dennoch sollten wesentlich mehr Noten eingetragen werden um das Ergebnis positiv mit korrekten Treffern zu beeinflussen.



### 9.1.3 Suchmaschinen mit Query by Humming Funktion

Im Verlauf werden 2 Portale vorgestellt und kleinere Tests zur Funktionalität und Präzision des jeweiligen QbH-Systems durchgeführt.

Portale: Musicline und Midomi

Testsongs: „A hard days night“ – The Beatles

“Like a prayer” – Madonna



#### 9.1.3.1 Musicline

Musicline ist die größte offizielle Musiksuchmaschine im deutschsprachigen Netz. Dem Nutzer stehen ca. 2 Mio. Songs zum Reinhören zur Verfügung, sowie die offiziellen Charts und Links zu Künstlern und Bands. Zudem bot musicline erstmalig im Web die *Query by Humming* Funktion an. Der Entwickler des QbH-Systems der Musiksuchmaschine „Musicline“ ist das *Fraunhofer-Institut für Digitale Medientechnologie*.

#### Suchemöglichkeiten

- Profisuche
- MelodieSuche
- Livesuche
- Genresuche

#### Metadaten

Über die „Profisuche“ gelangt der Nutzer zu der umfangreichen Metadatensuche.

The screenshot shows the 'Profisuche / Eingabe' (Advanced Search / Input) section of the Musicline website. It features several input fields and dropdown menus for filtering search results. The fields include: 'Künstler' (Artist), 'Album', 'Songtitel' (Song Title), 'Jahr der Veröffentlichung' (Year of Release) with a dropdown menu set to 'alle' (all), 'Monat der Veröffentlichung' (Month of Release) with a dropdown menu set to 'alle', 'Einträge pro Seite:' (Entries per page) with a dropdown menu set to '10', 'Komponist (nur Klassik)' (Composer (Classical only)), 'Bestellnr. (EANUPC)', 'Label', 'Vertrieb' (Distribution) with a dropdown menu set to 'Vertrieb auswählen', and 'Genre (maximal drei)' (Genre (maximal three)) with a text box containing 'Bitte unten Genre auswählen.' Below the genre text box are two buttons: 'Genres auswählen...' and 'Genre entfernen'. On the right side, there is a 'Tonträgerformat:' (Carrier format) section with a list of checkboxes: 'Alle' (checked), 'CD', 'CD-Maxi/Single', 'Vinyl', 'MC', 'DVD-Audio', 'SACD', 'DVD-Video', 'Video VHS', and 'Sonstige'. At the bottom left, there is a blue button labeled 'jetzt suchen' (search now).

Abbildung 27 : Metadatenuche von Musicline

Quelle: <http://www.musicline.de/de/prosearch>

## Genresuche

Die Genre von Musicline setzen sich zusammen aus: Pop, Rock, Dance & Electronic, Black Music, HipHop, Alternative, Metal, Klassik, Volksmusik, World, Schlager, Comedy, Jazz und Andere Genres. Das ist nur eine grobe Unterteilung. Über das Genrelexikon erfolgt eine feinere Unterteilung und Beschreibung.

Beispiel: Genre Rock

- Deutsch-Punk
- Deutsch-Rock
- Flower Power
- Folk Rock
- Glamrock
- Hamburger Schule
- Krautrock
- ...

Weiterhin können Künstlerbiographien eingesehen werden oder Neuveröffentlichungen. Die Suche erfolgt durch das schrittweise Auswählen der Kategorien.

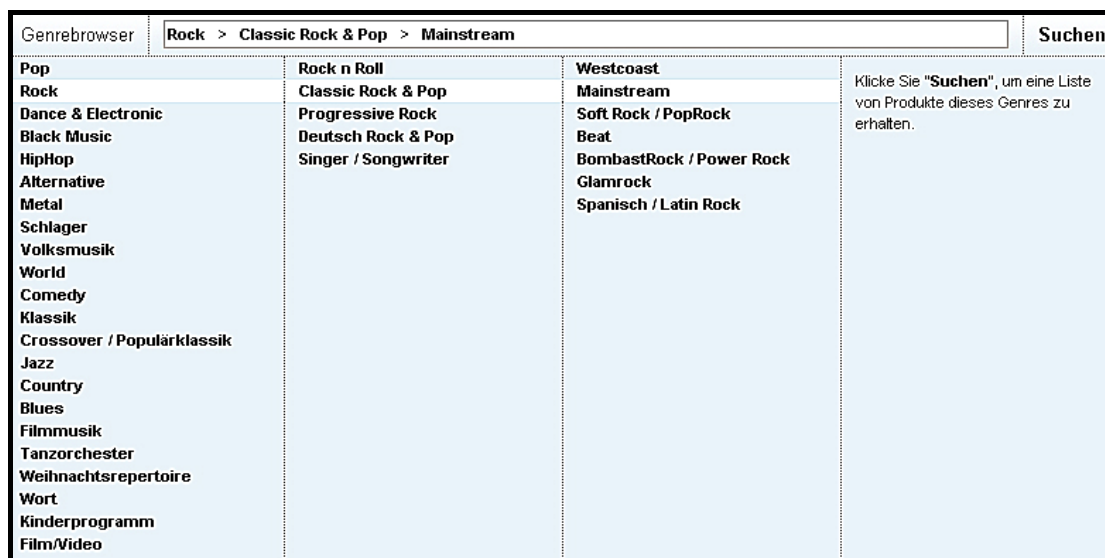


Abbildung 28 : Genresuche mit musicline

Quelle: <http://www.musicline.de/de/genresuche>

## Melodiesuche

Die Melodiesuche ist der Kernpunkt der Suchmaschine. Um diesen Service zunutzen, müssen verschiedene technische Voraussetzungen gegeben sein.

- Soundkarte mit Mikrofon-Eingang
- Angeschlossenes Mikrofon
- Java Runtime Environment 1.5

Das System unterstützt kein Pfeifen, kein leises Summen und keine Musik aus einer Stereoanlage. Des Weiteren versteht es keine Sprache und keine Songtexte. Die Nutzung der Suche ist kostenfrei.

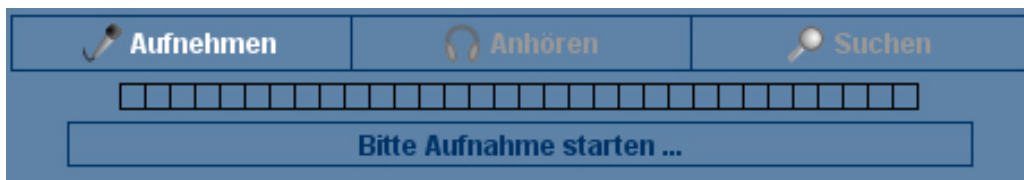


Abbildung 29 : Eingabe der Melodiesuche

### Test 1

Song: „A Hard days night“ – The Beatles

Gefundene Einträge (10)			
	Songtitel	Künstler	Score
1.	A Hard Day's Night	Beatles, The	<input type="range"/>
2.	Sexy	Westernhagen	<input type="range"/>
3.	I Can't Get Enough	Chyp-Notic	<input type="range"/>

Abbildung 30 : Ergebnis Test 1 - Musicline

### Test 2

Song: „Like a Prayer“ – Madonna

Gefundene Einträge (10)			
	Songtitel	Künstler	Score
1.	Like A Prayer	Madonna	<input type="range"/>
2.	Sweet Child O' Mind	Guns N' Roses	<input type="range"/>
3.	Liebe Pur	Cora	<input type="range"/>

Abbildung 31 : Ergebnis Test 2 - Musicline

### Ergebnis:

Beide Songs wurden gefunden und sind unter den „Top 3“. Was negativ auffiel war, dass die Lieder recht laut gesummt werden mussten, damit das System die Anfrage versteht. Ein weiterer Nachteil ist, dass sich nicht übermäßig viele Melodien in der Datenbank befinden.

### 9.1.3.2 Midomi



Midomi ist eine Musiksuchmaschine mit 2 Hauptsuchmethoden. Weiterhin können Gesänge aufgenommen und von anderen Nutzern bewertet werden.

#### Metadaten

Über ein einfaches Eingabefeld können Titel, Interpret, Band oder Album eingegeben werden. Danach wird eine Trefferliste aus den eingegebenen Informationen erzeugt. Midomi liefert dem Nutzer Informationen über den jeweiligen Künstler bzw. der jeweiligen Band.

Weiterhin kann in die Titel des Künstlers reingehört werden.

#### Sing or hum

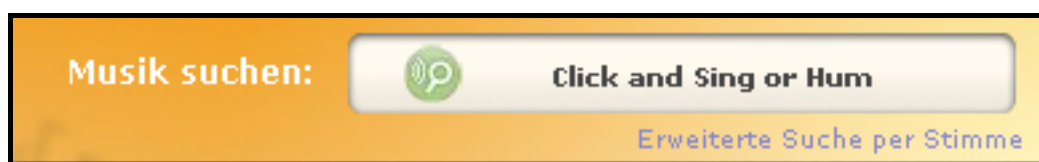


Abbildung 32 : Melodiesuche von Midomi

Quelle: <http://www.midomi.com>

Für die Suche wird ein Mikrofon benötigt. Über ein Flash Applet werden ca. 30 Sekunden des Gesangs aufgenommen und mit der internen Datenbank verglichen.

Die erweiterte Suche per Stimme bietet weiterhin die Funktion Genre und Sprache auszuwählen.

## Test 1

Song: „A Hard Days Night“ – The Beatles

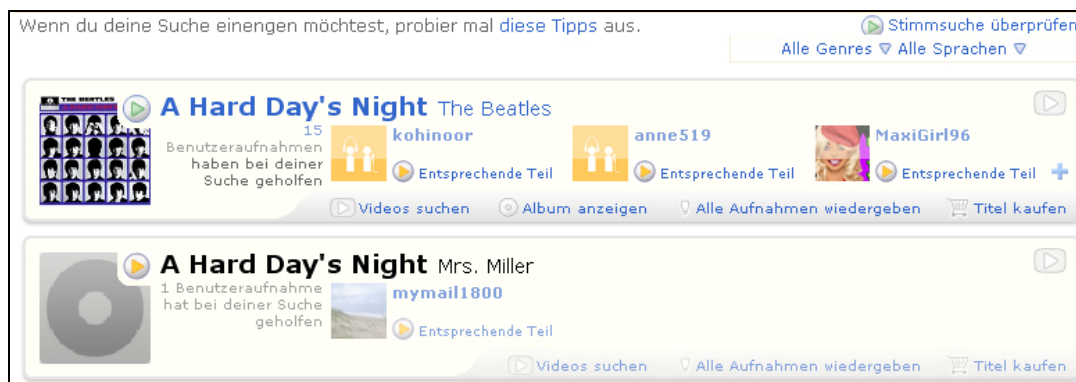


Abbildung 33 : Ergebnis Test 1 - Midomi

## Test 2

Song: "Like a Prayer" – Madonna



Abbildung 34 : Ergebnis Test 2 - Midomi

## Ergebnis

Beide Songs sind aufgeführt und auffindbar. Der Umfang der Melodien, die in der Datenbank gespeichert sind könnte größer sein.

## Endergebnis der Tests der QbH-Systeme

Generell funktioniert das System sehr gut, aber stößt beispielsweise an seine Grenzen bei „Dance“ Musik, welche keine Melodie aufweist. Weiterhin ist zwingend die Form des MIDI File mit seiner symbolischen Repräsentation notwendig.

## 9.2 Technologien von Musikempfehlungssystemen (englisch: recommendation systems)

Musikempfehlungssysteme erfreuen sich großer Beliebtheit, denn sie können auf Basis des Musikgeschmacks des Nutzers ein Profil aufbauen und ähnliche Musik anbieten. Die Systeme sammeln und verwalten personenbezogene Informationen, um Empfehlungen für den Nutzer zu erzeugen. Weiterhin bestimmen sie inwieweit ein Nutzer die Empfehlungsanfrage beeinflussen kann. Es gibt zahlreiche Verfahren für Empfehlungssysteme.<sup>144</sup>

### 9.2.1 Collaborative Filtering

Das *Collaborative Filtering* beschreibt eine Filtertechnik von Empfehlungssystemen durch Auswertung der „Weisheit von Vielen“ (engl. „Wisdom of the crowd“). Eine Anfrage wird gestellt und die Ergebnisse orientieren sich nach „Leute die diese Musik gehört haben, hörten auch“. Das Prinzip „beruht auf der Annahme, dass sich Menschen mit gleichen Interessen sich mit hoher Wahrscheinlichkeit auch für ähnliche Suchergebnisse interessieren“<sup>145</sup>. Die Vorschläge werden mit Hilfe von Algorithmen erstellt. Unterschieden wird CF in *user-based* und *item-based*. Mit dem *user-based* Prinzip werden ähnliche Nutzer (Ähnlichkeit im Sinne von beispielsweise einer Playlist) miteinander verglichen. Die Empfehlung basiert dann auf Songs, die der eine Nutzer hört, aber der Andere nicht kennt.

Dem Prinzip nach *item-based* folgen viele Online-Shops. Dem Kunden werden Produkte vorgeschlagen, die von anderen Kunden erworben wurden, die auch das aktuell betrachtete Produkt gekauft haben.<sup>146 147 148</sup>

---

<sup>144</sup> Vgl. Information Search and Retrieval.

<http://www.iicm.tugraz.at/cguetl/courses/isr/uearchive/uews2008/Ue04%20-%20Personalization%20in%20IR.pdf> [letzter Zugriff: 20.06.2009]

<sup>145</sup> Vgl. collaborative filtering.

[http://www.wissensnavigator.com/interface2/knowledge/applications/collaborative\\_filtering/index.htm](http://www.wissensnavigator.com/interface2/knowledge/applications/collaborative_filtering/index.htm) [letzter Zugriff: 19.06.2009]

<sup>146</sup> Vgl. Content-Based versus Collaborative

Filtering. <http://www.is-frankfurt.de/uploads/down417.pdf> [letzter Zugriff: 20.06.2009]

### 9.2.2 content based Retrieval

Im content based Retrieval ist die musikalische Ähnlichkeit von Bedeutung. Die Parameter für die Erstellung einer solchen Beschreibung sind beispielsweise das Tempo, die Instrumentation, die Harmonie, die Tonart oder Tonhöhe. Die extrahierten Informationen werden in Fingerprints zusammengefasst. Die Parameter werden durch Filterung und Transformation des Audiomaterials erschlossen.

Die Anfrage erfolgt mit einem bekannten Song des Nutzers. Die Datenbank überprüft die Ähnlichkeit der Merkmale mit den bereits gespeicherten Objekten und empfiehlt sie dem Nutzer.<sup>149 150 151</sup>

### 9.2.3 Tagging

Unter dem „Tagging“ bzw. „Social Tagging“ wird eine freie Vergabe von Schlagwörtern (Tags) für digitale Objekten verstanden, die von Nutzern durchgeführt wird. Es gibt viele Systeme die dieses Prinzip verfolgen. Im Musikbereich können sich solche Tags verschieden äußern, z.B. 80s, 90s, Rock. Das hat eine persönliche Organisation und Strukturierung zur Folge. Vorteile bestehen hierbei aber nicht ausschließlich für eine Person. Auch die Gemeinschaft des Systems profitiert von dem Prinzip. So können Sammlungen von Tags (Folksonomien) aufgrund der kollektiven Kategorisierung zur besseren Struktur der angebotenen Objekte beitragen. Beispielsweise bei dem Musikempfehlungssystem ist es möglich aufgrund vorhandener Folksonomien durch gleiche Tag-Vergabe bei unterschiedlichen Songs auf eine Ähnlichkeit zu schließen. Auch Nutzer können aufgrund ihrer vielfältigen Verwendung der Tags kategorisiert und miteinander verglichen

---

<sup>147</sup> Vgl. Fundamental Models and Algorithms for a Distributed Reputation System. <http://elib.uni-stuttgart.de/opus/volltexte/2008/3401/pdf/dissertation.pdf> [letzter Zugriff: 15.06.2009]

<sup>148</sup> Vgl. ASDL:Recommendersysteme. <http://www.is.informatik.uni-duisburg.de/wiki/index.php/ASDL:Recommendersysteme> [letzter Zugriff: 20.06.2009]

<sup>149</sup> ebenda

<sup>150</sup> ebenda

<sup>151</sup> Vgl. Content-Based versus Collaborative Filtering. <http://www.is-frankfurt.de/uploads/down417.pdf> [letzter Zugriff: 20.06.2009]



werden, was zu einer nutzerspezifischen Empfehlung anhand der Ähnlichkeit führt.<sup>152</sup>

#### 9.2.4 Hybride Systeme<sup>153</sup>

Unter „Hybriden Systemen“ wird die Kombination aus zwei oder mehrerer Verfahren verstanden. Aufgrund dessen können die Vorteile der verschiedenen Empfehlungstechniken gleichzeitig genutzt und eventuell Nachteile verringert werden. Häufig wird das Collaborative Filtering mit einer anderen Technik kombiniert um das „New User“<sup>154</sup> oder „New item“<sup>155</sup> Problem zu lösen.

### 9.3 Beispiele für Empfehlungssysteme

#### 9.3.1 Online-Shops

Viele Online-Musikshops wenden die Technik des Collaborative Filtering an, um den Nutzer zum Kauf des jeweiligen Produkts zu animieren.

##### 9.3.1.1 Amazon.de

Der Online-Shop *Amazon* bezeichnet sich selbst als „Social Commerce“ Versandhaus und bedient ein weiträumiges Spektrum an Artikeln, darunter auch Musik. Wenn ein Produkt erworben wird, stellt Amazon noch einige Kaufempfehlungen. Diese orientieren sich häufig an den Kauf- oder Suchverhalten anderer Kunden, die denselben Artikel erworben haben – Collaborative Filtering –item-based. Die Empfehlungen sind variabel und mit jeder neuen Bestellung und jedem neuen Suchvorgang werden sie

---

<sup>152</sup> Vgl. Nutzung kollektiven Wissens durch Social Tagging. <http://www.iwm-kmrc.de/www/de/projekte/projekt.html?name=SocialTagging&dispname=SocialTagging> [letzter Zugriff:20.06.2009]

<sup>153</sup> Methodenübersicht – siehe Anhang

<sup>154</sup> bei der Registrierung eines neuen Users sind noch keine Informationen bekannt, darum kann zu Beginn noch nichts empfohlen werden

<sup>155</sup> Item welches noch keine Bewertung durchlaufen hat, kann nur schwer empfohlen werden


aktualisiert. Sie erstrecken sich über die gesamte Produktpalette von amazon und sind nur für registrierte Nutzer sichtbar.

## Beispiel für Musikempfehlung

**Empfehlungen Musik**


- [Alternative](#)
- [Comedy & Kabarett](#)
- [Dance & Electronic](#)
- [Diverses](#)
- [Gothic & Wave](#)
- [Jazz & Blues](#)
- [Kindermusik & Hörspiele](#)
- [Klassik](#)
- [Klassik-Crossover](#)
- [Latin](#)
- [Metal & Hardrock](#)
- [Pop](#)
- [R&B & Soul](#)
- [Rap & HipHop](#)
- [Reggae & Ska](#)
- [Rock](#)
- [Rock & Pop aus Deutschland](#)
- [Schlager & Oldies](#)
- [Songwriter, Folk & Country](#)
- [Soundtracks & Musicals](#)
- [Weltmusik](#)

---

1.  **Back to Black incl. Valerie (Deluxe Edt.) [DOPPEL-CD]**  
von Amy Winehouse (November 30, 2007)  
Durchschnittliche Kundenbewertung: ★★★★★ (43)  
Auf Lager.  
**Preis: EUR 16,95**  
[53 Angebote](#) ab EUR 10,20

Diesen Artikel haben wir empfohlen, weil Sie u.a. **Frank** gekauft haben. ([Bitte ändern](#))

---

2.  **Take Off Your Pants and Jacket**  
von Blink 182 (Juni 19, 2001)  
Durchschnittliche Kundenbewertung: ★★★★★ (42)  
Auf Lager.  
**Preis: EUR 7,95**  
[36 Angebote](#) ab EUR 3,10

Diesen Artikel haben wir empfohlen, weil Sie **Blink 182** gekauft haben. ([Bitte ändern](#))

---


3.  **Only By The Night**  
von Kings of Leon (September 19, 2008)  
Durchschnittliche Kundenbewertung: ★★★★★ (72)  
Auf Lager.  
**Preis: EUR 12,45**  
[40 Angebote](#) ab EUR 9,47

Abbildung 35 : Ausschnitt des Empfehlungsbereiches von amazon

In der Abbildung ist zu erkennen, dass die Empfehlungen nach Musik in Genre unterteilt sind. Auch oben ist noch einmal der Verweis, dass diese Empfehlungen auf dem Kaufverhalten des Nutzers basieren.

Beispiel in der Abbildung: „Diesen Artikel haben wir empfohlen, weil Sie u.a. Ok Go gekauft haben“.

Durch das Bewertungssystem von amazon, können die Empfehlungen beeinflusst werden.

## Ungenauigkeit

Wenn ein Artikel bereits gekauft wurde, kann es vorkommen, dass amazon ihn dennoch empfiehlt. Das geschieht häufig bei Büchern, die es eventuell noch einmal in einer Taschenbuchausgabe gibt.

Weiterhin ist die Kategorisierung der Musik recht übersichtlich, leider aber stellenweise unverständlich.

Beispiel: Warum taucht bei den Empfehlungen in der Kategorie Metal & Hardrock die Band Blink 182 auf, die eigentlich zum Rock gehört?

### 9.3.2 Audioscrobbler

Ein Audioscrobbler ist ein Netzwerkdienst mit dem ein individuelles Musikprofil erstellt werden kann. Dabei muss der Nutzer sich zunächst (kostenlos) registrieren und eine Erweiterung seines favorisierten Musikabspielprogramms (z.B. Winamp) vornehmen. Danach werden die dort abgespielten Lieder „gescrobbelt“ und zu einem Profil aufgebaut. Seit der Fusion 2005 ist der Audioscrobbler bei Last.fm verfügbar.<sup>156</sup>

#### 9.3.2.1 Last.fm

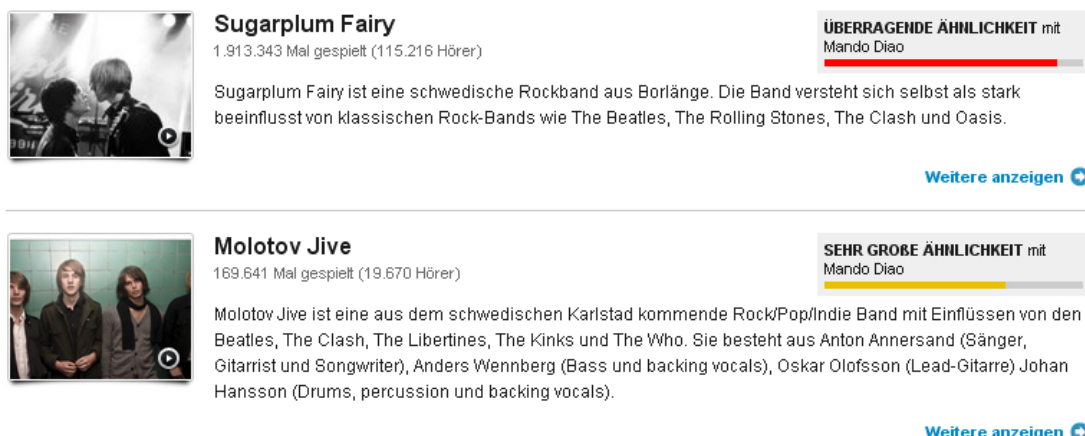
Das Internetradio Last.fm erstellt für den Nutzer Musikprofile auf Basis eines Audioscrobblers. Die abgespielten Lieder werden gesammelt und auf Ähnlichkeit überprüft. Das Resultat dieses Prozesses sind Musikempfehlungen, sowie Informationen über den jeweiligen Künstler. Das „scrobbeln“ von Musik bedeutet, dass die abgespielten Songs getrackt werden. Dabei wird auch die Dauer berücksichtigt, ob ein Lied vollständig gehört wird oder nur 20, 30 Sekunden.

Auf der Website selbst können verschiedene Zusatzinformationen über die Lieblingskünstler gewonnen werden.

Hierbei ist der Menüpunkt „Ähnliche Künstler“ interessant. Dieser zeigt eine Liste mit den Künstlern an, die eine starke Ähnlichkeit zu der Musik des soeben ausgewählten Künstlers aufweisen.

---

<sup>156</sup> Vgl. Audioscrobbler. <http://www.computerbase.de/lexikon/Audioscrobbler> [letzter Zugriff: 22.06.2009]



The screenshot shows two artist recommendation cards from Last.fm. The first card is for 'Sugarplum Fairy', a Swedish rock band from Borlänge, with 1,913,343 plays and 115,216 listeners. It features a red progress bar indicating 'ÜBERRAGENDE ÄHNLICHKEIT mit Mando Diao'. The second card is for 'Molotov Jive', a Swedish rock/pop/indie band from Karlstad, with 169,641 plays and 19,670 listeners. It features a yellow progress bar indicating 'SEHR GROßE ÄHNLICHKEIT mit Mando Diao'. Both cards include a 'Weitere anzeigen' button with a right-pointing arrow.

Abbildung 36 : „Ähnliche Künstler“ Funktion bei Last.fm

Auf der Abbildung ist rechts eine Leiste zu erkennen. Diese zeigt anhand verschiedener Farbvarianten, welche Künstler eine hohe Ähnlichkeit aufweisen. Anhand des Beispiels ist eine „überragende Ähnlichkeit“ mit der Band „Sugarplum Fairy“ vorhanden.

### 9.3.3 Grafische Musiksuche

Eine grafische Musiksuche basiert auf Ähnlichkeit und kann visuell anhand einer „cloud“ wiedergegeben werden.

#### 9.3.3.1 Music Map

Die Idee hinter Music Map ist recht simpel. Der Nutzer gibt den Namen seines favorisierten Künstlers ein und eine *Wolke* (cloud) wird geladen. Die Abstände zwischen den verschiedenen Ergebnissen symbolisieren dabei die Ähnlichkeit. Diese wird vermutlich durch das Genre bestimmt. Was sehr frustrierend bei dieser Darstellungsart sein kann ist, dass die Ergebnisse visuell überlagern oder sich immer wieder bewegen. Des Weiteren stellt sich auch hier öfter die Frage, warum gerade diese Band oder dieser Künstler empfohlen wird, obwohl (scheinbar) kein Zusammenhang ersichtlich ist.

### 9.3.4 Visuelle Darstellung des Musikgeschmacks

Musikvisualisierungen anhand des Musikgeschmackes des Nutzers sind eher eine Seltenheit bei den heutigen Portalen. Darunter gibt es nur wenige Ausnahmen.

#### 9.3.4.1 Musiccovery

Musiccovery ist ähnlich Last.fm ein Internetradio, beruht jedoch auf inhaltsbasierten Empfehlungen. Durch die Eingabe von Genre oder dem Veröffentlichungsjahr etc., werden die Interpreten ausgewählt. Des Weiteren visualisiert Musiccovery den Musikgeschmack eines Nutzers anhand seiner Lieblingssongs. Die Größe der Formen variiert dabei und gibt Aufschluss über die Beliebtheit. Die Farbe symbolisiert die Musikrichtung.

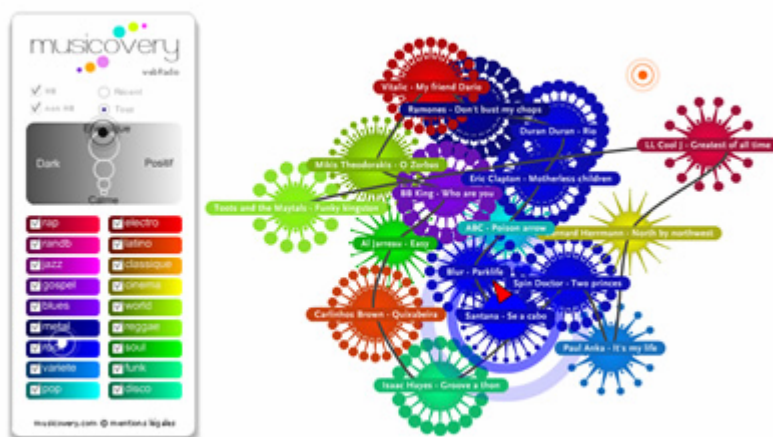


Abbildung 37 : Visualisierung des Musikgeschmacks von musiccovery

Quelle: <http://www.visualcomplexity.com/vc/index.cfm?all=yes>

## 9.4 Übersicht von einigen Portalen mit grober Funktionskategorisierung

### Legende

1	Musikempfehlung
2	Metadatenuche
3	Query by Humming
4	Query by Tapping
5	Parsons Code
6	Klaviatur Suche

Portal	Kurze Beschreibung	1	2	3	4	5	6
amazon	Onlineshop mit Empfehlungen basierend auf Collaborative Filtering – item-based	X					
Audiobaba	Suchmaschine und Musikempfehlungssystem basierend auf akustischer Ähnlichkeit, Collaborative Filtering - content- based, Erstellen von Playlists, Acoustic Matches (Titel eines Songs anklicken um Empfehlungen für ähnliche Musik zu erhalten), Hören von Samples	X	X				
Emusic	Online-Musikdienst, aktuell mehr als 4 Millionen Songs verfügbar, hauptsächlich Independent Music		X				
flyfi	Internetradio, eine Eingabe des Lieblingskünstlers ermöglicht den Aufbau eines eigenen Musikkanals mit ähnlichen Künstlern	X					
iLike	Musikservice, Collaborative Filtering, der Musikgeschmack anderer Nutzer kann anhand von abgespielten Songs und Künstlern eingesehen werden	X	X				
imeem	hochgeladene Musik von Benutzern wird gespeichert, Collaborative Filtering, besitzt umfangreichen Tracker, Tagging ist möglich, Player mit umfangreichen Funktionen, Musikbewertung	X	X				
inDiscover	Collaborative Filtering, Taggen beim Hochladen, spezialisiert auf Independent	X	X				
Last.fm	Internetradio, Audioscrobbing, Empfehlungen (Playlisten, Artists, Alben und Songs eines Benutzers und Userverhalten), eigene oder globale Tags, Anzeige von Musiksammlungen anderer Nutzer, Anzeige von Usern mit ähnlichem Musikgeschmack	X	X				
liveplasma	Darstellung des Musikgeschmacks	X					
Meemix	Internetradio, Player, Anlegen von Radiostationen möglich	X					
Midomi	Musiksuchmaschine, Query by Humming möglich	X	X	X			
MOG	Collaborative Filtering, Anzeige von Nutzern mit ähnlichen Musikgeschmack, Player,	X	X				

	Audioscrobbing						
mufin	Musiksuchmaschine, Collaborative Filtering - content based, Empfehlungen basieren auf ähnlichem Klang	X	X				
Musiclens	Musiksuchmaschine, Collaborative Filtering - content-based, Musikgeschmack wird mittels Schieberegler eingestellt	X					
Musicline	Musiksuchmaschine, 2,2 Millionen Songs zum Reinhören, Query-by.Humming möglich		X	X			
Musicoverly	Internetradio, Erstellen eines eigenen Radios, Collaborative Filtering- content based, Darstellung von Musikgeschmack	X	X				
musipedia	Musiksuchmaschine mit folgenden Funktionen: Query by whistle, Query by Tapping, Noten einzeichnen, Klaviatur Suche, Melodische Kontur			X	X	X	X
mystrands	Empfehlungssystem, Collaborative Filtering	X	X				
Pandora	Internetradio, Empfehlung nach ähnlichen Songs	X	X				
Qloud	Online-Dienst, Tagging durch User	X	X				
songtapper	Suchmaschine anhand von Query by Tapping				X		
the filter	Musikempfehlungssystem	X	X				
UpTo11	Musikempfehlungssystem, Collaborative Filtering, Analyse von Playlisten,	X	X				
ZuKool Music	Musikempfehlungssystem, Collaborative Filtering content-based	X					

Tabelle 7 : einfache Übersicht von Musikempfehlungssysteme und Musiksuchmaschinen

## Schluss / Zusammenfassung

Das Ziel der Arbeit war es, einen Einblick in die verschiedenen Aspekte des komplexen Bereichs des „*Music Information Retrieval*“ zu verschaffen. Von Bedeutung war besonders, wie ein Audiosignal anhand seiner charakteristischen Merkmale für eine Recherche einsetzbar ist. Zunächst wurden die verschiedenen grundlegenden Begrifflichkeiten aus der Musiklehre geklärt, mit den Basisdimensionen von Music Information Retrieval. Die folgenden Kapitel schilderten wie diese Merkmale extrahiert und in der Mustererkennung für Klassifikationen genutzt werden. Für Nutzer sollen Musikinformatoren schnell auffindbar sein, anhand von verschiedenen Suchmethoden, die nicht ausschließlich auf herkömmlichen Textretrieval basieren. Die Anwendungen sollten diese Methoden anhand von Beispielen mit kleineren Tests darstellen. Dabei zeigt sich, dass die gegenwärtigen Systeme einfach und präzise sind. Besonders die heutigen Musikempfehlungssysteme stechen durch ihre Vielfältigkeit aus der Menge heraus. Worauf in dieser Arbeit nicht eingegangen wurde, ist das Musikidentifikationssystem auch für Mobiltelefone entwickelt werden, die auf den „*Query by Humming*“ Systemen beruhen<sup>157</sup>. Diese haben bereits in den kleineren Tests dieser Arbeit sehr gut abgeschnitten. Die Systeme arbeiten sehr präzise und ermöglichen ein schnelles und korrektes Ergebnis. Sie zeigten sich selbst bei Stör- oder Nebengeräuschen äußerst tolerant. Auch die Audiosynchronisation wurde in der Arbeit kurz angerissen. Das Ziel einer hochauflösenden Synchronisation war bisher noch nicht möglich, obwohl bereits viele Ansätze entwickelt wurden, ähnlich wie bei der polyphonen Transkription. Inwieweit sich der Bereich Music Information Retrieval noch weiterentwickeln wird, lässt sich nur schwer erahnen, aber ohne diese Retrieval Systeme wäre eine Strukturierung und Nutzung des Segments des Online-Musikmarktes in der heutigen Zeit nur noch schwer vorstellbar.

---

<sup>157</sup> Vgl. Query by Humming Melodieerkennungssystem.  
[http://www.idmt.fraunhofer.de/de/presse\\_medien/download/produktinformation/09\\_qbh\\_de.pdf](http://www.idmt.fraunhofer.de/de/presse_medien/download/produktinformation/09_qbh_de.pdf) [letzter Zugriff:22.06.2009]

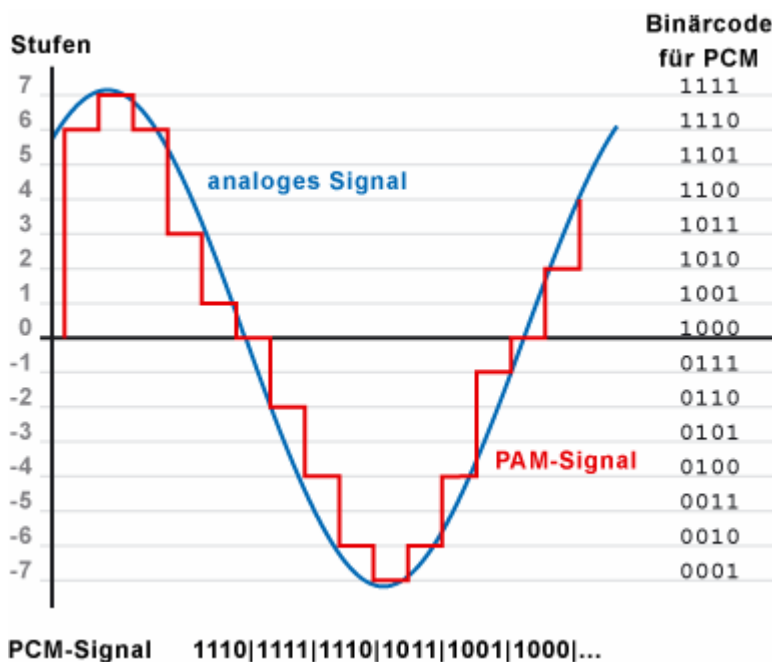


## Anhang

### 1. Schritte des Pulse-Code-Modulation Verfahrens<sup>158</sup>

#### Sample and Hold (Abtast- und Halteschaltung)

Das analoge Signal wird zunächst durch Abtastung in ein zeitdiskretes PAM Signal (Pulsamplitudenmodulation) umgewandelt. Anhand der Sample und Hold Stufe werden die abgetasteten Impulse abgeflacht und die ungleichmäßigen Impulse auf bestimmte Amplitudenstufen gebracht. Das Ausgangssignal ist dann ein zeitdiskretes, linearisiertes PAM-Signal in einer stufenartigen Form.



Quelle: <http://www.elektronik-kompodium.de/sites/kom/0312281.htm>

#### Quantisierung

Das Einteilen bzw. die Wertezuordnung des Amplitudenbereichs in Spannungsstufen wird als Quantisierung bezeichnet. Jedem analogen Wert

<sup>158</sup> Vgl. Quelle: <http://www.elektronik-kompodium.de/sites/kom/0312281.htm>

wird ein digitaler Wert zugeordnet. Die Quantisierung des PAM Signals erfolgt mit einem A/D Wandler.

## **Codierung**

Das Signal wird in binäre Datenworte gewandelt und den einzelnen Amplitudenwerten (Spannungsstufen) entsprechende Quantisierungsstufen und damit ein Binärcode zugeordnet. Die Übertragung der digitalen Signale erfolgt anhand des Binärcodes.

## **2. MPEG 7 Aufbau – Teilbereiche**

### **1. Systems**

- definiert Systemwerkzeuge zur Dokumentspeicherung,- und Übertragung und Synchronisation mit Nutzerdaten

### **2. Data Definition Language:**

- Definitionssprache zur Erzeugung von Beschreibungen
- Baut auf XML-Schema auf

### **3. Visual Descriptors**

- Grundlegende Beschreibungen von Merkmalen in Bezug auf Bilder und Videos
- Merkmale sind „*Low-Level*“ Eigenschaften wie Form, Bewegung, Farbe, Umrisse von Objekten, Bewegungsintensität

### **4. Audio Descriptors**

- Eigenschaften von Audiosignalen
- Low-Level-Eigenschaften auf variierender Abstraktionsebene  
Tonhöhe, Harmonie und Lautstärke

## 5. Multimedia Description Schemes

- Beschreibungsstrukturen
- Semantisch und struktureller Beschreibungsrahmen in den audiovisuelle *Descriptors* eingefügt werden

## 6. MPEG-7 Reference Software

- Beispiel-Implementierungen für Erzeugung und Verarbeitung der *PEG-7-Descriptor* und DSs

## 7. Conformance Testing

- Vorschriften für Konformitätstests der MPEG-7-Beschreibungen

## 3. Herleitung der diskreten Fourier Transformation

Die Formel der Fourier - Transformation bildet den Ausgangspunkt zur Herleitung der diskreten Fourier-Transformation. Das Ziel ist die Transformierung eines zeitkontinuierlichen Signals.

$$x(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt$$

Die Abtastung dient dabei der Überführung des zeitkontinuierlichen- in ein zeitdiskretes Signal.

Änderung der Formel:

- $X(t)$  wird durch  $x(nT)$  ersetzt – seine Abtastwerte
- Differential – Abtastintervall  $T$
- Integral - Summenzeichen

$$x_x(f) = \sum_{n=-\infty}^{\infty} x(nT) e^{-j2\pi nT f}$$

Aus den unendlichen Abtastwerten wird eine endliche Zahl  $n$  heraus geschnitten. Durch Weglassen des Faktors  $T$  erhält man das abgetastete und gefensterete Signal.

$$x_s w(f) = \sum_{n=-\infty}^{n-1} x(nT) e^{-j2\pi n \frac{f}{f_s}}$$

Die Funktion ist periodisch und besitzt an den  $N$  Frequenzstellen linear unabhängige Frequenzwerte, daher werden die  $N$  Frequenzstellen folgendermaßen ausgewertet.

$$x_s w\left(k \frac{f_s}{N}\right) = \sum_{n=-\infty}^{N-1} x(nT) e^{-j2\pi n \frac{k f_s}{N}}, k = 0, 1, 2, \dots, N-1$$

Im letzten Schritt der Herleitung wird  $\text{sw}$  und der Faktor  $\frac{f_s}{N}$  nicht angezeigt und die Abtastwerte sind nun  $x(n)$ .

$$x_k = \sum_{n=-\infty}^{N-1} x(n) e^{-jk \frac{2\pi}{N}} \quad \text{Diskrete Fourier-Transformation}$$

#### 4. spektrale und zeitdiskrete Merkmale<sup>159</sup>

Spectral Centroid

- bestimmt den Schwerpunkt des Signals im Spektrum

$$\text{Formel: } SC(b) = \frac{\sum_{f=0}^{F-1} f \cdot S_b(f)^2}{\sum_{f=0}^{F-1} S_b(f)^2}$$

- $F$  = Anzahl der Frequenzen im Spektrum

<sup>159</sup> Vgl. <http://archiv.tu-chemnitz.de/pub/2008/0018/data/Klausung-Genreklassifikation.pdf>

- $S_b$  = Anzahl der Frequenzen des Zeitbereichs
- $B, S_b(f)$  = Amplitudenwert der Frequenz  $f$  im Spektrum  $S_b$

### Spectral Flux

- Betrachtung von mehreren Spektren
- Erkennung von Veränderungen der Blöcke

$$\text{Formel: } SF = \frac{1}{(B-1)(F-1)} \sum_{b=1}^{B-1} \sum_{f=1}^{F-1} (\log(S_b(f) + d) - \log(S_{b-1}(f) + d))^2$$

- $B$  = Gesamtzahl der Blöcke
- $d$  = verhindert Überläufe

### Spectral Rolloff

- bestimmt die Höhe eines bestimmten Energieanteils im Spektrum

$$\text{Formel: } SRF = \max \left( h \left| \sum_{f=0}^h S_b(f) \leq TH \cdot \sum_{f=0}^{F-1} S_b(f)^2 \right. \right)$$

- $TH$  = Menge der Energie, soll unterhalb der gesuchten Frequenz  $h$  stehen

### Zero Crossing Rate

- Anzahl der Durchgänge des Signals durch den Nullpunkt

$$\text{Formel: } ZCR = \frac{1}{N-1} \sum_{n=0}^{N-1} |\text{sgn}(x_n) - \text{sgn}(x_{n-1})|$$

- $N$  = Anzahl der Samples im Zeitfenster der Berechnung
- $x_i$  = Samplewert an der Stelle  $n$

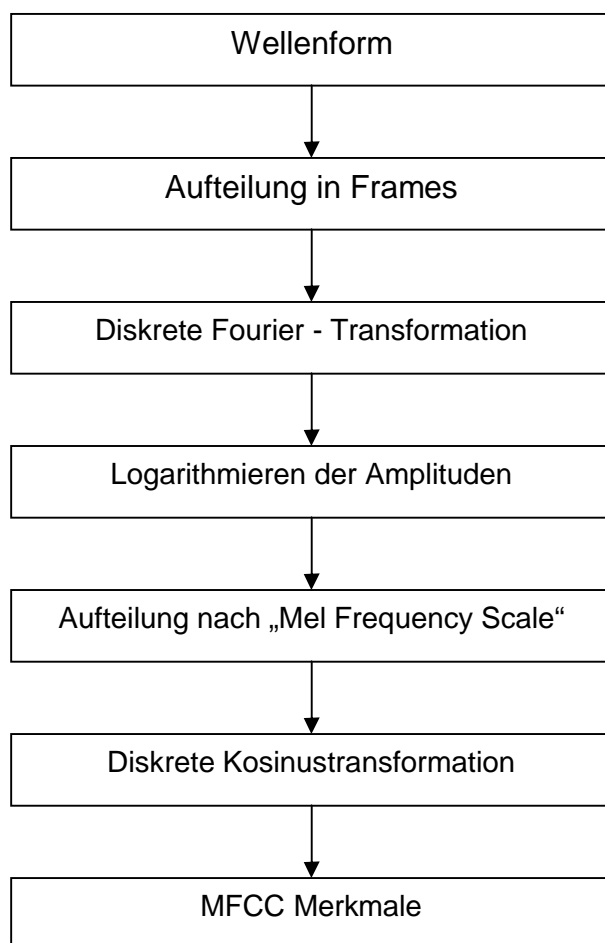
- $\text{sgn}(x)$  Variante der Signum-Funktion

### Root Mean Square

- Energie eines Signals

Formel: 
$$RMS = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x_n^2}$$

### 5. Mel-Frequency Cepstral Coefficients – Schritte<sup>160</sup>



---

<sup>160</sup> <http://archiv.tu-chemnitz.de/pub/2008/0018/data/Klausing-Genreklassifikation.pdf>

Die Abbildung zeigt die Schritte der Berechnung von MFCC Merkmalen. Mit den Merkmalen ist eine aussagekräftige Charakterisierung mit nur wenigen Werten eines Audiosignals möglich

#### Prozess

- Berechnung der Frames anhand der DFT
- Umwandlung der Amplituden Skala in eine logarithmische Skala
- Phaseninformation des Spektrums wird verworfen
- Reduzierung der einzelnen Amplitudenwerte auf eine geringere Menge von Bereichen
- Bereiche sind nach „Mel Frequency Scale“ aufgeteilt
- Werte besitzen hohe Korrelation untereinander → Einsatz der Hauptkomponentenanalyse (PCA)
- PCA sehr rechenintensiv → Approximation mit Diskreter Kosinustransformation
- Ergebnis: Merkmalsvektor mit vorgegebener Anzahl von Dimensionen

## 6. Schritte zur Erstellung eines Chromagramm

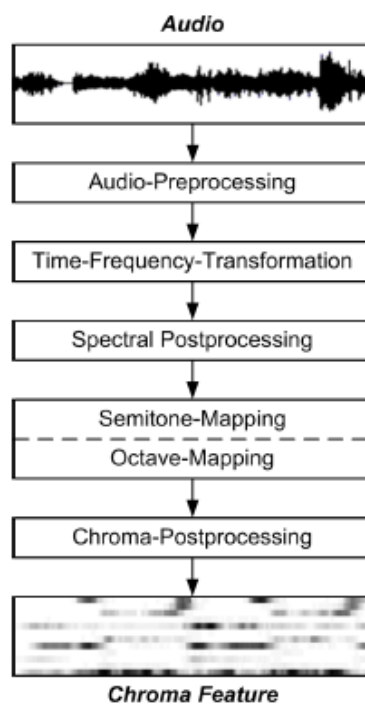


Abbildung : Schritte von einem Audio-Signal zu einem Chromagramm

Quelle: "Evaluation and Comparison of Audio Chroma Feature Extraction Methods"

## 7. Spektrogramm

Spektrogramme sind grafische Darstellungen des Spektrums im Zeitverlauf.

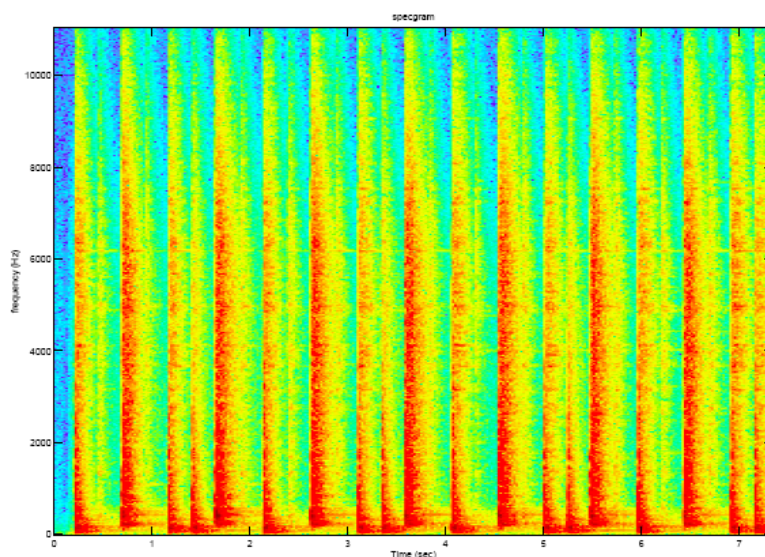


Abbildung: Spektrogramm eines einfachen drumsounds



Quelle: [http://www.hobbyist.de/dsp/fae\\_anzengruber.pdf](http://www.hobbyist.de/dsp/fae_anzengruber.pdf)

Je wärmer die Farben desto höher die Amplitude im Frequenzbereich.  
Je vielfältiger die Instrumentierung und je komplexer das Signal, desto schwerer ist es klare Strukturen zu erkennen und herauszulesen.

Rote Stellen = hohe Energiewerte

Blauen Stellen= niedrige Energiewerte

## 8. Kombinationsmöglichkeiten – Hybride Systeme

<b>Weighted</b>	Die Bewertungen von verschiedenen Recommendation-Techniken werden kombiniert, um eine einzelne Bewertung zu erhalten
<b>Switching</b>	Das System wechselt je nach Situation die Recommendation-Technik
<b>Mixed</b>	Ergebnisse verschiedener Recommendation-Systeme werden gleichzeitig angezeigt
<b>Feature Combination</b>	Features von verschiedenen Recommender-Datenquellen werden zu einem einzigen Algorithmus zusammengefaßt
<b>Cascade</b>	Ein Recommender verfeinert das Ergebnis eines anderen Systems
<b>Feature augmentation</b>	Output einer Methode ist die Eingabe einer anderen
<b>Meta-level</b>	Das von einem Recommender gelernte Modell ist die Eingabe eines anderen Systems

Quelle: <http://www.is.informatik.uni-duisburg.de/wiki/index.php/ASDL:Recommendersysteme>

## Quellenverzeichnis

### *Printquellen*

Stock, Wolfgang G. (2006): Information Retrieval. Suchen und Finden von Informationen. – München; Wien: Oldenbourg Wissenschaftsverlag

Batke, Johann M. (2006): Untersuchung von Melodiesuchsystemen sowie von Verfahren zu ihrer Funktionsprüfung. –Göttingen: Cuvillier Verlag

Downie, Stephen J.: Music Information Retrieval (Kapitel 7), In: Annual Review of Information Science and Technology (2002) 37, S. 295-340.- Information Today

Wendemuth, Andreas (2004): Grundlagen der stochastischen Sprachverarbeitung -Oldenbourg

Müller, Meinhard (2007): Information Retrieval for Music and Motion. – Berlin: Springer Verlag

Meyer, Martin (2008): Signalverarbeitung: Analoge und digitale Signale, Systeme und Filter.- Wiesbaden : Vieweg+Teubner Verlag

Plato, Robert(2004): Übungsbuch zur Numerischen Mathematik: Aufgaben, Lösungen und Anwendungen.- Wiesbaden : Vieweg+Teubner Verlag

Holst, Imogen(1992): Das ABC der Musik: Grundbegriffe, Harmonik, Formen, Instrumente. - Ditzingen: Reclam Verlag

Schmitt, Ingo (2005): Ähnlichkeitssuche in Multimedia-Datenbanken: Retrieval, Suchalgorithmen und Anfragebehandlung.-München: Oldenburg Verlag

### *Internet-Quellen*

Müller, Meinard; Kurth, Frank; Clausen, Michael (2006): Aktuelle Aspekte des Music Information Retrieval.

Online verfügbar unter <http://www.datenbank-spektrum.de/pdf/dbs-18-24.pdf> [Letzter Zugriff: 11.06.2009]

Das Elektronik Kompendium: PCM – Pulsmodulation. Online verfügbar unter <http://www.elektronik-kompendium.de/sites/kom/0312281.htm> [Letzter Zugriff: 15.06.2009]

Höldrich, Robert: Algorithmen in Akustik und Computermusik. Online verfügbar unter <http://www.iem.at/lehre/skripten/>[Letzter Zugriff: 11.05.2009]

Müller, Stephan: Expectation Maximization Algorithmus und Modellschätzung. Online verfügbar unter <http://www-mmdb.iai.uni-bonn.de/lehre/proprak0304/mueller.pdf> [letzter Zugriff: 01.06.2009]

Skowronek, Janto; McKinney, Martin; van de Par, Steven; Breebaart, Jeroen (2008): Anwendungsbeispiele für Musikanalyse-Algorithmen. Online verfügbar unter <http://www.jeroenbreebaart.com/papers/daga/daga2008.pdf> [letzter Zugriff: 11.06.2009]

Shaykhit, Dmytro: Inhaltsbasierte Musiksuche: Konzepte und Anwendungen. Online verfügbar unter [http://www.medien.ifi.lmu.de/fileadmin/mimuc/hs\\_ws0506/papers/MIR.pdf](http://www.medien.ifi.lmu.de/fileadmin/mimuc/hs_ws0506/papers/MIR.pdf) [Letzter Zugriff: 11.06.2009]

Ziehe, Katrin (2005): Melodie-Suche n einem Archiv von historischen Notenhandschriften. Online verfügbar unter [http://www.enotehistory.de/db/SA\\_KatrinZiehe\\_2006.pdf](http://www.enotehistory.de/db/SA_KatrinZiehe_2006.pdf) [letzter Zugriff: 22.06.2009]

Veltkamp, Remco C. ; Wiering, Frans; Typke, Rainer: Content Based Music Retrieval. Online verfügbar unter <http://www.cs.uu.nl/groups/MG/multimedia/publications/art/mme-music.pdf> [Letzter Zugriff: 11.06.2009]

Tzanetakis, George; Essl, Georg; Cook, Perry: Automatic Musical Genre Classification Of Audio Signals. Online verfügbar unter <http://ismir2001.ismir.net/pdf/tzanetakis.pdf> [Letzter Zugriff: 11.06.2009]

Ewert, Sebastian (2007): Effiziente Methoden zur hochauflösenden Musiksynchronisation. Online verfügbar unter [http://www.mpi-inf.mpg.de/~mmueller/publications/2008\\_Diplomarbeit\\_EwertSebastian\\_MusikSync.pdf](http://www.mpi-inf.mpg.de/~mmueller/publications/2008_Diplomarbeit_EwertSebastian_MusikSync.pdf) [letzter Zugriff: 12.06.2009]

Thiel, Sebastian(2005): Herleitung der diskreten Fourier-Transformation. Online verfügbar unter <http://www.uni-koblenz.de/~physik/informatik/DSV/DFT.pdf> [letzter Zugriff: 12.06.2009]

Tzanetakis, George; Ermolinskyi, Andreye; Cook, Perry: Beyond the Query-By-Example Paradigm: New Query Interfaces for Music Information Retrieval. Online verfügbar unter <http://www.cs.berkeley.edu/~andreye/papers/icmc02.pdf> [letzter Zugriff: 24.06.2009]

Ellis, Daniel P.W.: CLASSIFYING MUSIC AUDIO WITH TIMBRAL AND CHROMA FEATURES. Online verfügbar unter <http://www.ee.columbia.edu/~dpwe/pubs/Ellis07-timbchroma.pdf> [letzter Zugriff: 24.06.2009]

- Lukas, Anzengruber (2005): Projektbericht Feature Extraction Audio A. Online verfügbar unter [http://www.hobbyist.de/dsp/fae\\_anzengruber.pdf](http://www.hobbyist.de/dsp/fae_anzengruber.pdf) [letzter Zugriff: 24.06.2009]
- Bindreiter, Stefan (2006): Audio gesteuerte Animationen. Online verfügbar unter <http://work.dabindi.net/masterthesis/Stefan%20Bindreiter.pdf> [letzter Zugriff: 24.06.2009]
- Jaenecke, Joachim(2008): Musikinformation und Musikdokumentation. Online verfügbar unter [http://www.miz.org/static\\_de/themenportale/einfuehrungstexte\\_pdf/08\\_MedienRecherche/jaenecke.pdf](http://www.miz.org/static_de/themenportale/einfuehrungstexte_pdf/08_MedienRecherche/jaenecke.pdf) [letzter Zugriff: 24.06.2009]
- Müller, Meinard; Kurth, Frank; Clausen, Michael: Automatisierte Annotation von Audiodaten mittels Synchronisationstechniken. Online verfügbar unter <http://subs.emis.de/LNI/Proceedings/Proceedings68/GI-Proceedings.68-11.pdf> [letzter Zugriff: 24.06.2009]
- Deibert, Sina: Query-by-humming im Internet. Online verfügbar unter [http://pi3.informatik.uni-mannheim.de/downloads/hauptstudium/seminare/papers/Ausarbeitung\\_QueryByHumming.pdf](http://pi3.informatik.uni-mannheim.de/downloads/hauptstudium/seminare/papers/Ausarbeitung_QueryByHumming.pdf) [letzter Zugriff: 24.06.2009]
- Chon, SongHui (2007): A Study of Synchronization of Audio Data with Symbolic Data. Online verfügbar unter <http://ccrma.stanford.edu/~shchon/projects/Music254/project.pdf> [letzter Zugriff: 24.06.2009]
- Hu, Ning; Dannenberg, Roger B.; Tzanetakis, George: Polyphonic Audio Matching and Alignment for Music Retrieval. Online verfügbar unter <http://www.cs.cmu.edu/~rbd/papers/waspaa03alignment.pdf> [letzter Zugriff: 24.06.2009]
- Macrae, Robert: Linking music-related information and audio data. Online verfügbar unter <http://138.37.35.209/digitalmusic/people/robm/PhD%20Draft.pdf> [letzter Zugriff: 24.06.2009]
- Bello, Juan P.; Pickens, Jeremy (2005): A Robust Mid-level Representation for Harmonic Content in Music Signals. Online verfügbar unter <http://www.elec.qmul.ac.uk/people/juan/Documents/Bello-ISMIR-2005.pdf> [letzter Zugriff: 24.06.2009]
- Mzyk, Michael (2005): Extraction of Rhythmical Features of Audio Signals. Online verfügbar unter <http://cvpr.uni-muenster.de/teaching/ws05/seminarWS05/downloads/ExtractionRhythmicalFeaturesAudioSignals-Ausarbeitung.pdf> [letzter Zugriff: 24.06.2009]
- Marolt, Matija (2006): A Mid-level Melody-based Representation for Calculating Audio Similarity. Online verfügbar unter

<http://lgm.fri.uni-lj.si/~matic/clanki/ismir2006.clanek.pdf> [letzter Zugriff: 24.06.2009]

Kreutzer, Sebastian (2008): Beiträge zum Entwurf und Vergleich von robusten Audiomerkmalen für das Musikretrieval. Online verfügbar unter [http://www.mpi-inf.mpg.de/~mmueller/publications/2008\\_Diplomarbeit\\_KreuzerSebastian\\_AudioMatching.pdf](http://www.mpi-inf.mpg.de/~mmueller/publications/2008_Diplomarbeit_KreuzerSebastian_AudioMatching.pdf) [letzter Zugriff: 24.06.2009]

Bortey, Daniel (2005): Die Verwendung des MPEG-7 Standards in der Szenen Interpretation. Online verfügbar unter <http://www.sts.tu-harburg.de/pw-and-m-theses/2005/bort05.pdf> [letzter Zugriff: 22.06.2009]

Klausing, Tilo; Seifert, Frank (2007): Robuste Genre-Klassifikation. Online verfügbar unter <http://archiv.tu-chemnitz.de/pub/2008/0018/data/Klausing-Genreklassifikation.pdf> [letzter Zugriff: 22.06.2009]

Habich, Dirk; Hinneburg, Alexander: Ähnlichkeitssuche in Musik-Datenbanken mit Hilfe von Visualisierungen. Online verfügbar unter <http://doesen0.informatik.uni-leipzig.de/proceedings/paper/91.pdf> [letzter Zugriff: 23.06.2009]

Hartmann, Julian (2005): Music Information Retrieval Automatische Genre-Klassifikation. Online verfügbar unter <http://cvpr.uni-muenster.de/teaching/ss05/blockseminarSS05/downloads/Music-Ausarbeitung.pdf> [letzter Zugriff: 23.06.2009]

Siegemund, Jan: Hauptkomponentenanalyse. Online verfügbar unter <http://www-mmdb.iai.uni-bonn.de/lehre/proprak0304/siegemund.pdf> [letzter Zugriff: 23.06.2009]

Rohdenburg, Thomas (2003): Klassifikation von Audio-Signalen. Online verfügbar unter [http://medi.uni-oldenburg.de/members/thomas/SoundClassification\\_Main\\_Color.pdf](http://medi.uni-oldenburg.de/members/thomas/SoundClassification_Main_Color.pdf) [letzter Zugriff: 23.06.2009]

Annesi, Paolo; Basili, Roberto; Gitto, Raffaele; Moschitti, Alessandro: Audio Feature Engineering for Automatic Music Genre Classification. Online verfügbar unter <http://riao.free.fr/papers/122.pdf> [letzter Zugriff: 23.06.2009]

Burred, Juan José; Lerch, Alexander (2003): A HIERARCHICAL APPROACH TO AUTOMATIC MUSICAL GENRE CLASSIFICATION. Online verfügbar unter <http://www.elec.qmul.ac.uk/dafx03/proceedings/pdfs/dafx06.pdf> [letzter Zugriff: 23.06.2009]

Lemmé, Thomas: Audio Retrieval. Online verfügbar unter [http://www.lemme.at/files/GWA\\_8\\_Audioretrieval.pdf](http://www.lemme.at/files/GWA_8_Audioretrieval.pdf) [letzter Zugriff: 23.06.2009]

Bröder, Aline (2005); Standards zur Modellierung von

Metadaten für digitale Musik. Online verfügbar unter [http://is-frankfurt.de/veranstaltung/Groffmann\\_SS05/Thema%205\\_Broeder\\_dl.pdf](http://is-frankfurt.de/veranstaltung/Groffmann_SS05/Thema%205_Broeder_dl.pdf) [letzter Zugriff: 23.06.2009]

Ilic, Manoela (2004): MUSIK-METADATEN. Online verfügbar unter <http://www-lehre.inf.uos.de/~milic/Coxi/Musik-Metadaten.pdf> [letzter Zugriff: 23.06.2009]

Lerch, Alexander: Ein Ansatz zur automatischen Erkennung der Tonart in Musikdateien. Online verfügbar unter [http://www.zplane.de/uploads/files/Lerch\\_TMT04\\_Ein\\_Ansatz\\_zur\\_automatischen\\_Erkennung\\_der\\_Tonart\\_in\\_Musikdateien.pdf](http://www.zplane.de/uploads/files/Lerch_TMT04_Ein_Ansatz_zur_automatischen_Erkennung_der_Tonart_in_Musikdateien.pdf) [letzter Zugriff: 20.06.2009]

Barwich, Axel; Holzmann, Alexander; Mieves, Dominic: Theoretische Grundlagen und eine Einführung in Messverfahren der Raumakustik. Online verfügbar unter <http://barwich.googlepages.com/akumessv05.pdf> [letzter Zugriff: 10.06.2009]

ZeM College GbR: MIDI Kompendium. [http://www.zem-college.de/midi/mc\\_allg.htm](http://www.zem-college.de/midi/mc_allg.htm) [letzter Zugriff: 16.05.2009]

Becker, Marcel: Audio Information Retrieval (2): Klassifizierung von Geräuschen. Online verfügbar unter <http://www-mmdb.iai.uni-bonn.de/lehre/proprak0304/becker.pdf> [letzter Zugriff: 24.06.2009]

Eppelt, Roland M.(2002): Analyse von Audiosignalen nach musikalischen Kriterien. Online verfügbar unter <http://roland.eppelt.de/da/Diplomarbeit.pdf> [letzter Zugriff: 20.06.2009]

Stommel, M.; Kuhnert, K.-D.: Übersicht Lernverfahren. Online verfügbar unter <http://www.uni-siegen.de/fb12/ezls/lehre/dbv2/seminar/05-kurzbeschreibunglernverfahren.pdf> [letzter Zugriff: 24.06.2009]

Skowronek, Janto; McKinney, Martin; van de Par, Steven; Breebaart, Jeroen (2008): Anwendungsbeispiele für Musikanalyse-Algorithmen. Online verfügbar unter <http://www.jeroenbreebaart.com/papers/daga/daga2008.pdf> [letzter Zugriff: 20.06.2009]

König, Wolfgang: Content-Based versus Collaborative Filtering. Online verfügbar unter <http://www.is-frankfurt.de/uploads/down417.pdf> [letzter Zugriff: 20.06.2009]

Paul, Alexandra (2005): Überblick über die methodischen Ansätze von Musikempfehlungssystemen. Online verfügbar unter [http://is-frankfurt.de/veranstaltung/Groffmann\\_SS05/Thema%207\\_Paul\\_dl.pdf](http://is-frankfurt.de/veranstaltung/Groffmann_SS05/Thema%207_Paul_dl.pdf) [letzter Zugriff: 20.06.2009]

Peters, Geoffrey; Anthony, Caroline; Schwartz, Michael (2005): Song Search and Retrieval by Tapping. Online verfügbar unter

<http://www.aaai.org/Papers/AAAI/2005/ISD05-014.pdf> [letzter Zugriff: 23.06.2009]

Kepper, Johannes: Codierungsformen von Musik. Online verfügbar unter [http://www.adwmainz.de/fileadmin/adwmainz/MuKo\\_Veranstaltungen/S2-Digitale\\_Medien/kepper.pdf](http://www.adwmainz.de/fileadmin/adwmainz/MuKo_Veranstaltungen/S2-Digitale_Medien/kepper.pdf) [letzter Zugriff: 01.06.2009]

Fraunhofer Institut für Digitale Medientechnologie: Query by Humming Melodieerkennungssystem. Online verfügbar unter [http://www.idmt.fraunhofer.de/de/presse\\_medien/download/produktinformation/09\\_qbh\\_de.pdf](http://www.idmt.fraunhofer.de/de/presse_medien/download/produktinformation/09_qbh_de.pdf) [letzter Zugriff: 22.06.2009]

### *Weitere Quellen*

Stein, Michael; Schubert, Benjamin M.; Gruhne, Matthias; Gatzsche, Gabriel; Mehnert, Markus (2009): Evaluation and Comparison of Audio Chroma Feature Extraction Methods.

Gouyon, Fabien; Dixon, Simon; Pampalk, Elias; Widme, Gerhard: Evaluating rhythmic descriptors for musical genre classification.

Tzanetakis, George: Tempo Extraction using Beat Histograms.

Tzanetakis, George; Cook, Perry(2002); Musical Genre Classification of Audio Signals.

Dunker, Peter; Nowak, Stefanie; Begau, André; Lanz, Cornelia: Content-based Mood Classification for Photos and Music.

Celma Herrada, Oscar(2009): Music recommendation and discovery in the long tail.

### *Homepages*

Homepage des Online-Shops amazon. Online verfügbar unter <http://www.amazon.de/>. [letzter Zugriff: 15.06.2009]

Homepage von audiobaba. Online verfügbar unter <http://www.audiobaba.com/>. [letzter Zugriff: 15.06.2009]

Homepage von Emusic. Online verfügbar unter <http://www.emusic.com/>. [letzter Zugriff: 15.06.2009]

Homepage von flyfi. Online verfügbar unter <http://www.flyfi.com/> [letzter Zugriff: 15.06.2009]

Homepage von iLike. Online verfügbar unter <http://www.ilike.com/> [letzter Zugriff: 15.06.2009]

Homepage von imeem. Online verfügbar unter <http://www.imeem.com/> [letzter Zugriff: 15.06.2009]

Homepage von inDiscover. Online verfügbar unter <http://www.indiscover.net/> [letzter Zugriff: 15.06.2009]

Homepage von Last.fm. Online verfügbar unter <http://www.last.fm/> [letzter Zugriff: 15.06.2009]

Homepage von liveplasma. Online verfügbar unter <http://www.musicplasma.com/> [letzter Zugriff: 15.06.2009]

Homepage von Meemix. Online verfügbar unter <http://www.meemix.com/app/> [letzter Zugriff: 15.06.2009]

Homepage von midomi. Online verfügbar unter <http://www.midomi.com/> [letzter Zugriff: 15.06.2009]

Homepage von musuma. Online verfügbar unter <http://www.musuma.de/> [letzter Zugriff: 15.06.2009]

Homepage von MOG. Online verfügbar unter <http://mog.com/> [letzter Zugriff: 15.06.2009]

Homepage von mufin. Online verfügbar unter <http://www.mufin.com/> [letzter Zugriff: 15.06.2009]

Homepage von musiclens. Online verfügbar unter <http://finetunes.musiclens.de/> [letzter Zugriff: 15.06.2009]

Homepage von musicline. Online verfügbar unter <http://www.musicline.de/> [letzter Zugriff: 15.06.2009]

Homepage von musicoverly. Online verfügbar unter <http://www.musicoverly.com/> [letzter Zugriff: 15.06.2009]

Homepage von musipedia. Online verfügbar unter <http://de.musipedia.org/melodiesuchmaschine.0.html?L=2> [letzter Zugriff: 15.06.2009]

Homepage von mystrands. Online verfügbar unter <http://www.mystrands.com/> [letzter Zugriff: 15.06.2009]

Homepage von Pandora. Online verfügbar unter <http://www.pandora.com/> [letzter Zugriff: 15.06.2009]

Homepage von Qloud. Online verfügbar unter <http://www.qloud.com/> [letzter Zugriff: 15.06.2009]



Homepage von songtapper. Online verfügbar unter <http://www.bored.com/songtapper/> [letzter Zugriff: 15.06.2009]

Homepage von the filter. Online verfügbar unter <http://www.thefilter.com/> [letzter Zugriff: 15.06.2009]

Homepage von UpTo11. Online verfügbar unter <http://www.upto11.net/> [letzter Zugriff: 15.06.2009]

Homepage von ZuKool Music. Online verfügbar unter <http://www.zukool.com/> [letzter Zugriff: 15.06.2009]

Homepage der Software für Musiknotation „CSound“. Online verfügbar unter <http://www.csounds.com/about> [letzter Zugriff: 12.06.2009]

Homepage von Recordare: „Music XML Definition“. Online verfügbar unter <http://www.recordare.com/xml.html> [letzter Zugriff: 20.06.2009]

SyncPlayer Homepage. Online verfügbar unter <http://audentify.iai.uni-bonn.de/synchome/> [letzter Zugriff: 20.06.2009]

Homepage der Suchmaschine midomi. Online verfügbar unter <http://www.midomi.com/> [letzter Zugriff: 20.06.2009]

Homepage der Suchmaschine Musipedia. Online verfügbar unter <http://de.musipedia.org/melodiesuchmaschine.0.html?L=2> [letzter Zugriff: 20.06.2009]

Homepage von musicoverly. Online verfügbar unter [1 http://www.musicoverly.com/index3.php?ct=de](http://www.musicoverly.com/index3.php?ct=de) [letzter Zugriff: 20.06.2009]