# Low Level Descriptoren

Anne Scheidler

# Aufbau des Vortrags

- LLD Kategorien
- Signaldarstellung
- Zeitbasierte Signaldarstellung und Merkmalsextraktion
- Transformation zwischen Signaldarstellungen
- Frequenzbasierte Signaldarstellung und Merkmalsextraktion
- MfCC (Begriffsdefinition und Berechnung)
- Rhythmus Content Feature

# Kategorien der Low Level Descriptoren (nach Tzanetakis)

- Timbral Texture Features (Klangfarbe)
  Sub Features , MFCC, Spectral Rolloff, Spectral Centroid, RMS

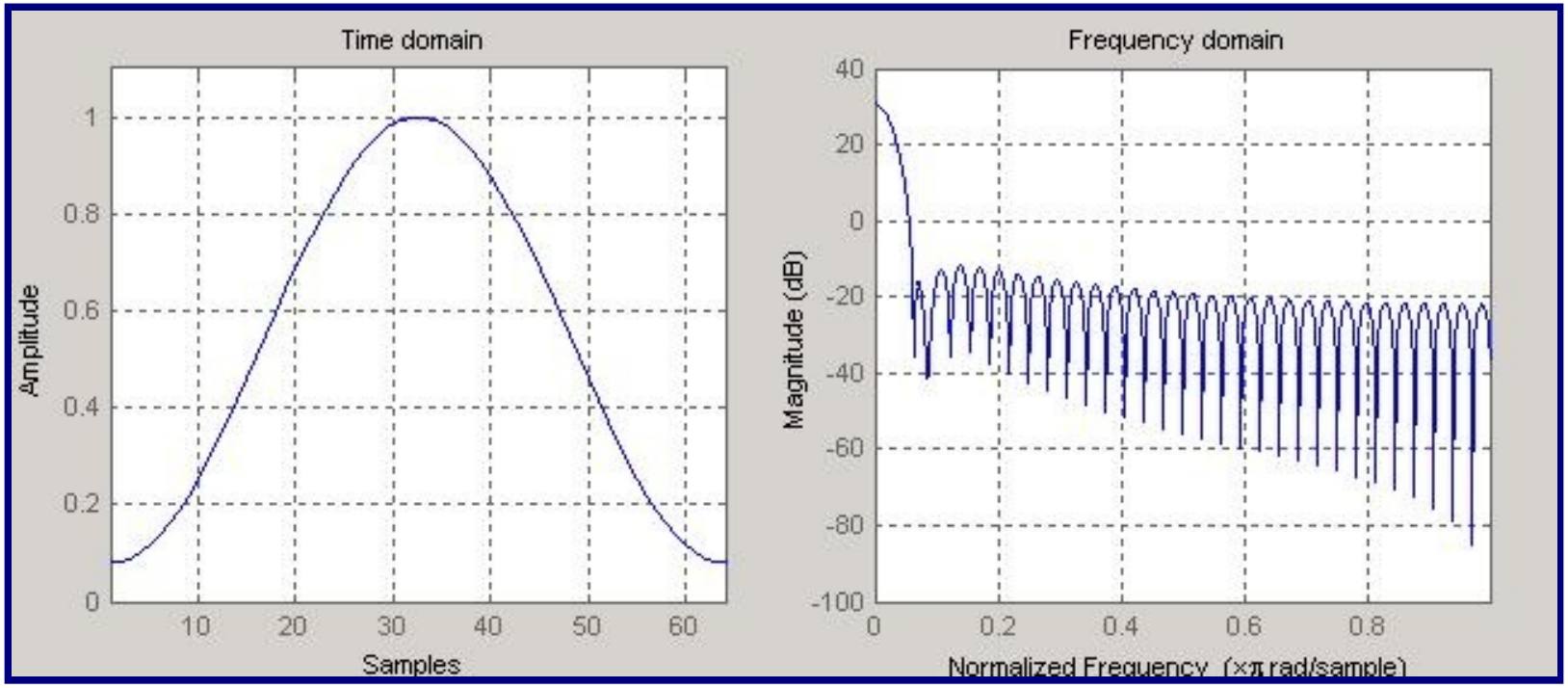
- Pitch Content Features (Tonhöhe)
  Pitch Histogramme Subfeatures u.a. Amplitudendifferenz

- Rhytmic Content Features (Rhytmus)
  Beat Histogramme u.a Merkmalsextraktion wie SUM

# Signaldarstellung

Grundlegende Darstellungsmöglichkeit von Signalen:

- Zeitbasiert  (Amplitude über Zeit), d.h. die Informationen liegen als Zeitreihen vor.
- Frequenzbasiert (Energiemenge über Frequenz in db) als Frequenzanteil, um Aussagen über das Frequenzspektrum machen zu können.
- Aus unterschiedlichen Darstellung lassen sich unterschiedliche Merkmale extrahieren.




- Merkmale werden nicht über das ganze Signal berechnet sondern über kleine Fensterausschnitte, die sich überlappen, so genannte Analysefenster.
- Wie viele Analysefenster und wie groß sollten diese gewählt sein ?
- Mehrere Analysefenster bilden das Texturenfenster.

# Sampling

- Abtastung: Die Registrierung von Messwerten zu diskreten, meist äquidistanten Zeitpunkten. Aus einem zeitkontinuierlichen Signal wird so ein zeitdiskretes Signal gewonnen.
- Abtastrate: Die Abtastrate bezeichnet die Rate, mit der Signalwerte aus einem kontinuierlichen Signal entnommen werden.
- Analoge Signale: Wenn eine wert- und zeitkontinuierliche Zuordnung von einer physikalischen Messgröße zu einer anderen (z.B. Temperatur) vorgenommen wird.
- Digitale Signale: Gegenteil von analogen Signalen. Digitale Signale liegen als wert- und zeitdiskrete Zahlenfolgen vor.

# Shannon-Nyquist Abtasttheorem

- Wenn man ein analoges Signal in ein digitales Signal umwandeln möchte, kann man die dafür notwendige Abtastrate durch das Abtasttheorem von Nyquist und Shannon bestimmen. Dabei gilt: Die Abtastrate muss mindestens doppelt so hoch wie die höchste im Signal vorhandene Frequenz sein.
- Untere Grenzfrequenz = 0

$$f_{\mathrm{abtast}} > 2\, f_{\mathrm{max}}$$

- Allgemein

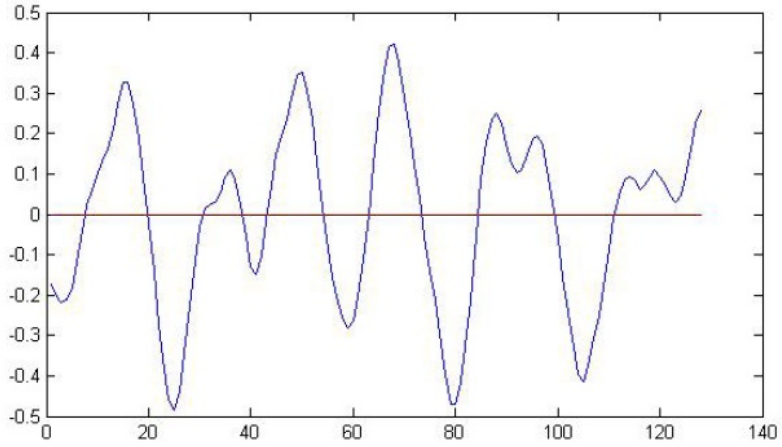
$$f_{\mathrm{abtast}} > 2(\, f_{\mathrm{obereGrenzfrequenz}} - f_{\mathrm{untereGrenzfrequenz}})$$

- Der Kehrwert der Abtastrate entspricht dem zeitlichen Abstand zwischen zwei Abtastungen.

# Zeitbasierte LLD - Zero Crossing Rate

Charakterisiert die Häufigkeit des Vorzeichenwechsels im Signal, d.h. wie oft überquert das Signal die 0-Amplitude

$$Z_t = \frac{1}{2}\sum_{n=1}^{N}\left|sign(x[n]) - sign(x[n-1])\right|$$



# Zeitbasierte LLD – Silent Ratio

Anteil der Stille einer Periode, d.h. die Anzahl der Messwerte einer Periode mit Amplitude = 0 ( quasi )

Gebräuchliche Schwellenwerte zur Berechnung:

1. Amplitudenwert unterhalb dessen Stille angenommen wird

2. Mindestanzahl von direkt aufeinander folgenden Messwerten, die Kriterium 1 erfüllen, um eine Stilleperiode zu bilden

# Darstellungstransformation

Um Aussagen über das Frequenzspektrum zu machen, muss von der zeitbasierten in die frequenzbasierte Darstellung gewechselt werden.

Hierfür stehen mehrere Möglichkeiten zur Verfügung:

- Eindimensionale Fourier Transformationen ( DFT, FFT )
- Wavelet Transformationen
- Cosinus Transformationen

# Fourier Transformationen FT

Jean-Baptiste-Joseph Fourier:
Man kann Funktionen durch die Summe von Sinus- und Cosinusfunktionen darstellen.

Grundlagen:

**Definition:** Sei $\mathbb{C}$ der Körper der komplexen Zahlen. Ein Element $w \in \mathbb{C}$ heißt $n$-te Einheitswurzel, wenn $w^n = 1$ ist.

$w$ heißt primitive $n$-te Einheitswurzel, wenn $w^n = 1$ ist, aber $w^k \neq 1$ für alle $k \in \{1, ..., n-1\}$.

Direkte Betrachtung der diskreten Fourier Transformation, da wir durch das digitalisierte Signal von einem diskreten Definitionsbereich ausgehen.

# Diskrete Fourier Transformation DFT

**Definition:** Sei $n \in \mathbb{N}$ und $w$ primitive $n$-te Einheitswurzel in $\mathbb{C}$. Die $n \times n$-Matrix $F$ mit

$$F_{i,j} = w^{i \cdot j}$$

für alle $i, j \in \{0, \ldots, n-1\}$, heißt Fouriermatrix.

Die lineare Abbildung $f : \mathbb{C}^n \rightarrow \mathbb{C}^n$ mit

$$f(a) = a \cdot F$$

für alle (Zeilen-)vektoren $a \in \mathbb{C}^n$ heißt Diskrete Fouriertransformation (DFT).

**Komplexität O ( $n^2$ )**

Anmerkung:

Die $k$-te Komponente des Ergebnisvektors $f(a)$ ergibt sich somit als Produkt des Vektors $a$ mit der $k$-ten Spalte der Fouriermatrix:

$$f(a)_k = \sum_{i=0,\ldots,n-1} a_i \cdot w^{i \cdot k}.$$

# Schnelle Fourier Transformation FFT

Die FFT ist ein Algorithmus von Cooley und Tukey der das selbe Ergebnis wie die DFT hat die Laufzeit aber auf  O ( n log(n)).

Idee:

- Voraussetzung: Anzahl der Abtastpunkte Zweierpotenz, somit Länge des Eingangsvektor z.B. 1,2, 4,8 , 16 usw. ( Radix 2 FFT)

- Divide And Conquer: Problem der Größe n in zwei Hälften der Größe n/2 Teilen. Der Messwertevektor wird nach geraden, ungeraden Indizes in Teilvektoren gesplittet.

- Die Ergebnisse der beiden Hälften werden dann zusammengeführt.

- Rekursive Anwendung der Grundidee impliziert die Laufzeit.

# Magnitude

- Bei der Transformation von der Zeit- in die Frequenz-Ebene entstehen aus Amplitude und Zeit eine reelle und eine imaginäre Zahl, die für die Berechnung der Magnitude dienen.

- Die Formel für diese Berechnung lautet :

$$Mag = \sqrt{R^2 + I^2}$$

(Mag = Magnitude / R = Reelle Zahl / I = Imaginäre Zahl)

Die Magnitude wird in db angegeben und ist auf der „Y-Achse des Spektrums“ abzulesen

# Spektrumbasierte LLD - Rolloff Grenzfrequenz

- Gibt an bis zu welcher Frequenz sich ein bestimmter Prozentsatz a des Gesamtspektrums aufsummieren lässt ( Cook/Tzanetakis a = 85)

$$\sum_{n=1}^{R_t} M_t[n] = \frac{\alpha}{100} \sum_{n=1}^{N} M_t[n]$$

Mt[n] Magnitudenwert des Spektrums an der Stelle n

- Höherer  Rolloff = höhere / stärkere Frequenzen

# Spektrumbasierte LLD - Spectral Centroid

Spectral Centroid, auch Brightness genannt, basierend auf der DFT, umgangssprachlich Klangfarbe

$$C_t = \frac{\sum_{n=1}^{N} M_t[n] \cdot n}{\sum_{n=1}^{N} M_t[n]}$$

Mt[n] Magnitudenwert des Spektrums an der Stelle n

C ist höher wenn bei gleich bleibender Tonhöhe viele Harmonien (d.h. Vielfaches der Grundfrequenz) vorkommen.

Ziel: Ermittlung welche Frequenz dominiert, also den Schwerpunkt des Spektrums

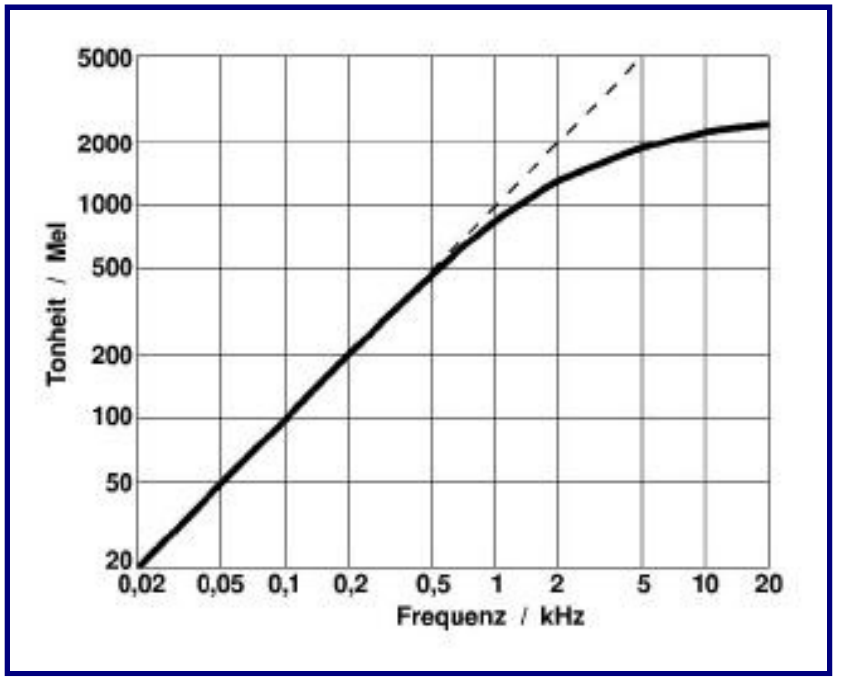
# Mel Frequency Cepstral Coefficients (MFCC)

Was ist Mel ?

Mel Z ist die Einheit für die psychoakustischen Größe der Toneinheit und beschreibt die wahrgenommene Tonhöhe. ( 1937 )

Basis für die Mel Skala ist der Ton C mit f = 131 Hert / Z = 131 mel

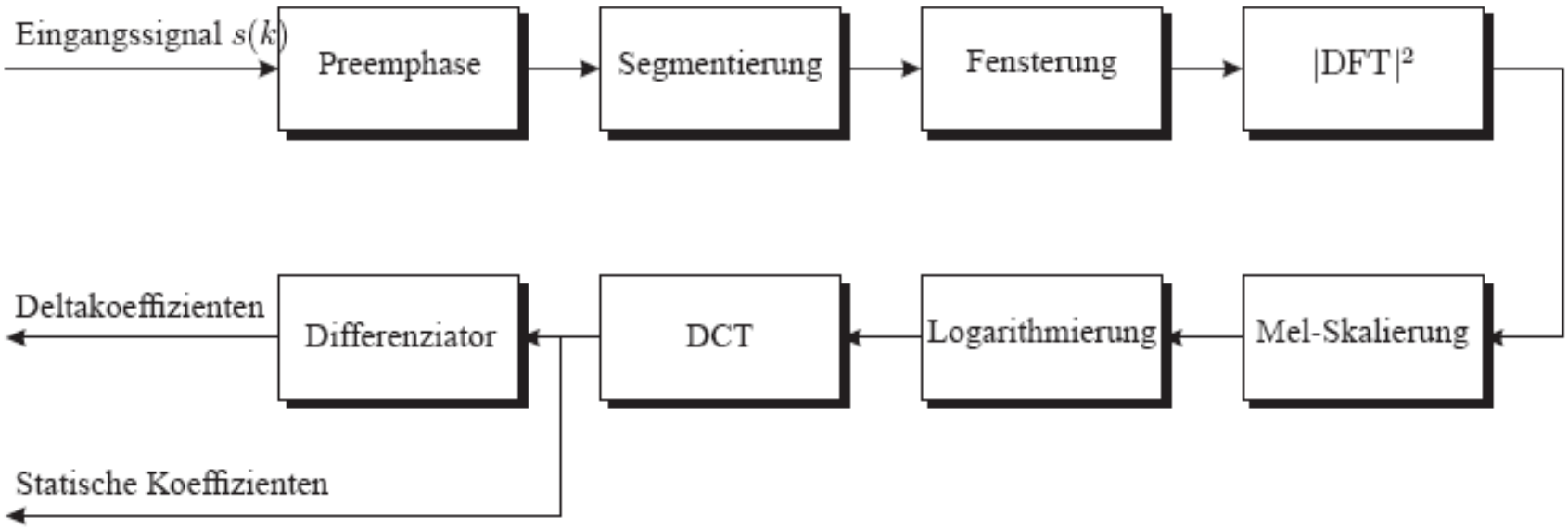
Mit Hilfe psychoakustischer Versuche kann so eine Tonheitsskala bestimmt werden (doppelt so hoch wahrgenommener Ton kriegt doppelten mel Wert)



# MFCC – elementare Begriffe

- Verfahren kommt ursprünglich aus der Spracherkennung
- Mel Cepstrum Koeffizient ist ein geeigneter Satz von Merkmalen zur automatischen Erkennung für die Musik
- Cepstrum = „Spektrum des Spektrums“, d.h. DCT eines logarithmischen Spektrums (ursprünglich für seismische Echos erfunden)
- Sprache verwendet 13 Koeffizienten
- Musik benötigt für effektive Ergebnisse nur die ersten 5 Koeffizienten (Tzanetakis/Cook)

# Ablauf MFCC


Eingangssignal $s(k)$
Preemphase
Segmentierung
Fensterung
$|\mathrm{DFT}|^2$
Mel-Skalierung
Logarithmierung
DCT
Differenziator
Deltakoeffizienten
Statische Koeffizienten

# Mathematik MFCC 1/4

1. PreEmphasis: Bewusste Verstärkung des Signals zur Anhebung der hohen Frequenzen. Da das Signal bereits diskret vorliegt, geschieht dieses mit Hilfe eines diskreten Filters, z.B. ein Fir-Filter.

3. Framing: Aufteilen des Signals in Bereich konstanter Größe, auf denen dann separat weitere Berechnungen durchgeführt werden. Die typischen Fenstergrößen sind 25 – 50 ms. (ms = Millisekunden)

5. Fensterung: Auf jeden dieser Bereich (Frame) wird nun die Fensterfunktion angewendet. Typisch bei MFCC ist das Hamming-Fenster, wobei generell die Auswahl der Fensterfunktion von der zugrunde liegenden Anwendung abhängt. Anwenden heißt, dass jeder Frame mit der Hamming-Fenster Funktion ausmultipliziert wird.

# Mathematik MFCC 2/4

1. DFT: Auf jeden Frame wird nun die DFT angewendet (2. Beachte die Framegröße, sie muss eine zweier Potenz sein, wenn FFT gewünscht wird). Das Signal liegt nun als Frequenzspektrum vor.

3. Mel Filter: Nun wird der Mel Filter angewendet, d.h. es wird eine Filterbank bestehend aus Dreiecksfiltern erstellt. Die Filter der Filterbank schwanken je nach Anwendung. Durch die Dreiecksfilter werden die Frequenzbänder (Frequenzbereiche) zusammengefasst, also die Komplexität reduziert, das Ergebnis ist das so genannte Mel-Spektrum.

5. Logarithmieren: Auf dem Mel Spektrum erfolgen nun weitere Berechnungen, deren Ergebnis Mel Koeffizienten (auch Kanalenergien) sind. Durch Logarithmieren dieser erfolgt eine weitere Zusammenfassung. Genauere Analysen sind jedoch erst nach Schritt 7. möglich.

# Mathematik MFCC 3/4

1. Cosinus Transformation: Das logarithmierte Mel Spektrum wird nun die diskrete Kosinus Transformation unterworfen. Im Unterschied zur FT werden nicht alle Frequenzen gleichmäßig behandelt, was zu einer Redundanzreduktion führt. Veranschaulicht wird das Spektrum geglättet, d.h. hochfrequente kleine Änderungen zwischen Werten entfernt. Ergebnis: Koeffizienten unkorreliert (Folge : eigene Aussagekraft einzelner Koeffizienten).

Alternativ Dekorrelation (Reduzieren der Redundanz): Principal Component Analyse, Karhunen-Loeve-Transformation

Definition Korrelation:

Kovarianz = Zusammenhang von Variablen, K = 0 unkorelliert, sonst korreliert

# Mathematik MFCC 4/4

Ergebnis: N Koeffizienten, zusammengefasst zu einem Merkmalsvektor X, die aus dem Mel-Frequenz-Spektrum gewonnen wurden und daher auch Mel-Frequency-Cepstral-Coefficients heißen.

Für die Merkmalsextraktion aus Musikdaten konnte gezeigt werden, dass die 5 ersten Werte aus dem gewonnenen Vektor ausreichend zur „Klassifizierung“ sind.

In der Sprecherkennung sind ersten 13 Koeffizienten aus dem gewonnen Vektor der gebräuchliche Wert.

Niedrige Cepstralwerte in Vektoren weisen auf hohe Frequenzen hin.

# Rhytmic Content Features

- Wavelet Transformation: Funktionen können auch durch die Summe von anderen Funktionen (Basisfunktionen) dargestellt werden.

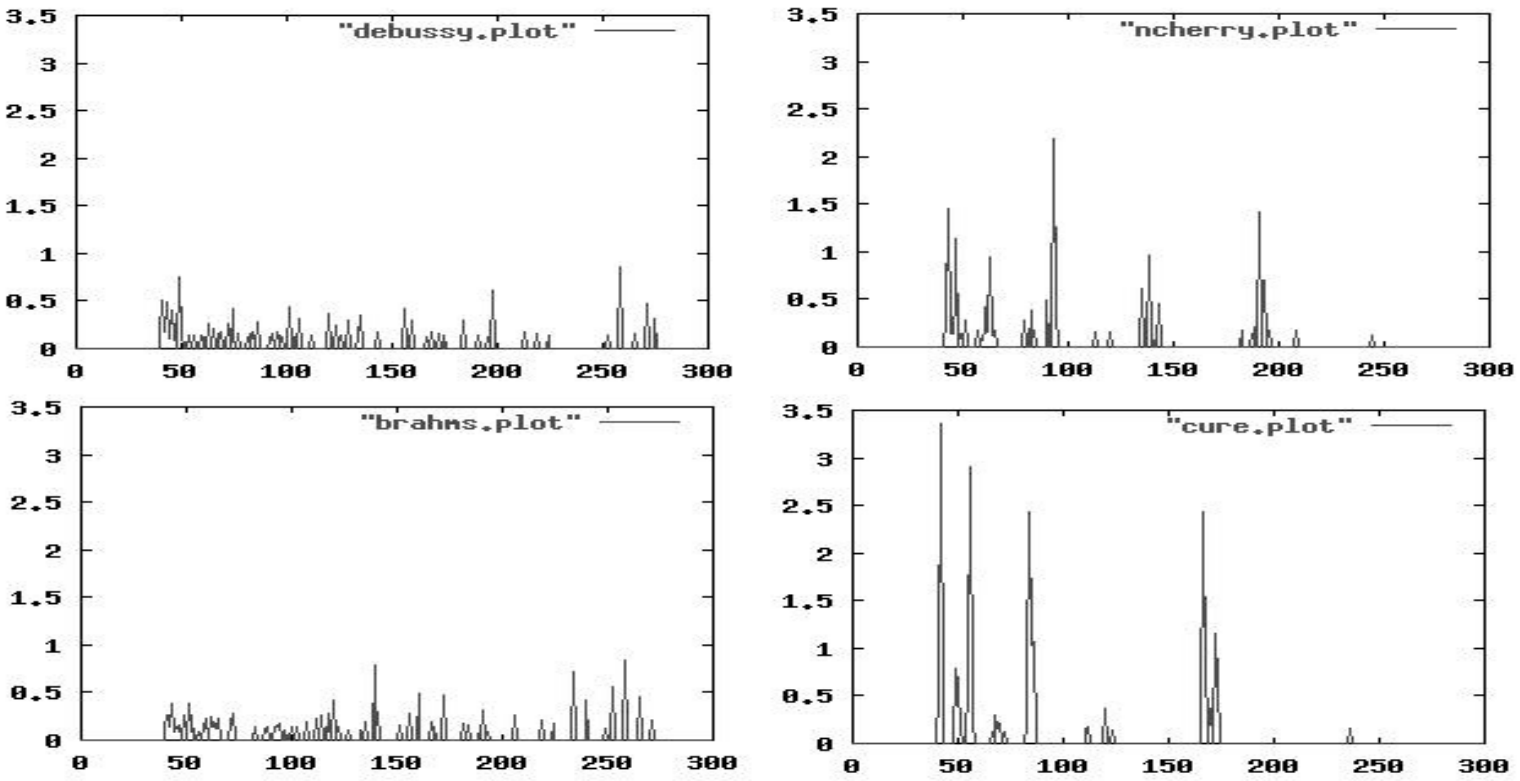
  Als Basisfunktion kann jede orthogonale Funktion genommen werden, für die gilt:

$$\int_{-\infty}^{\infty} h(t)\,dt = 0$$

Daher auch die Bezeichnung Wavelet engl. Wave = Welle

- Rhythmische Regelmäßigkeiten suchen
- Beat Histogramm
- Struktur + Stärke des Merkmals Rhythmus ablesen

# Beispiel Beat Histogramm



Vielen Dank für Ihre Aufmerksamkeit