Content-Based Audio Management

Music Information Retrieval

Jake Atwell, Philip Jakob

Agenda

- Einleitung/Motivation
- Muscle Fish Verfahren
- Baum-Verfahren
- Vergleich
- Fazit

Agenda

- **Einleitung/Motivation**
- Muscle Fish Verfahren
- Baum-Verfahren
- Vergleich
- Fazit

Einleitung

- Aktuelle Methoden für Audiosuche
  - subjektives Text
- Inhaltsbasiertes suchen und vergleichen
  - Die Signal selbst analysieren
  - PCA, RCA, Tempo-based, CLAS
- Mögliche Klassifizierungen
  - Simile
    - bellende Hunde
  - Akustische Eigenschaften
    - Helligkeit, Lautheit, Tonhöhe, usw.
  - Subjektive Eigenschaften
    - entspannt, aggressiv, usw.
  - Onomatopöie
    - BAM!

Einleitung

Signal

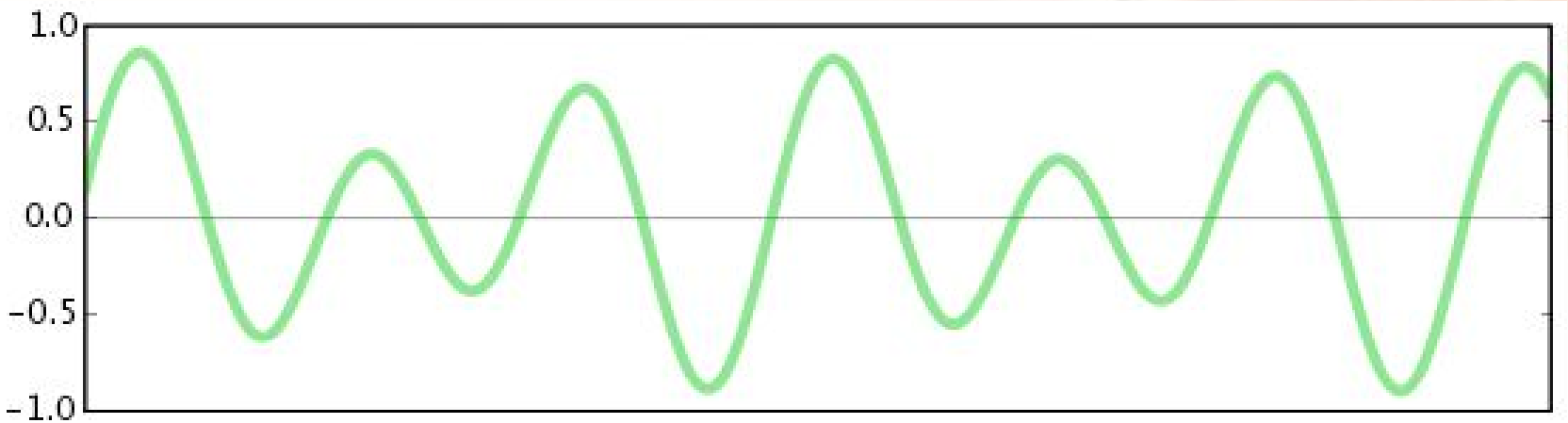


Image adapted from
http://manual.audacityteam.org/man/digital_audio.html

Einleitung

Signal

Samplingrate

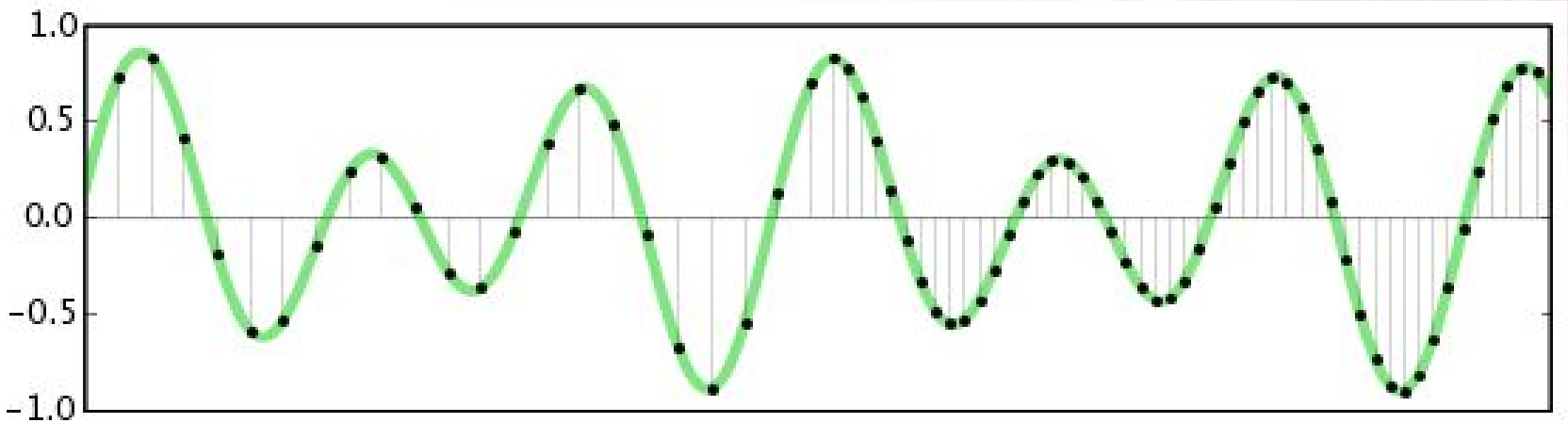


Image adapted from
http://manual.audacityteam.org/man/digital_audio.html

Einleitung

Signal Samplingrate Samplingstiefe

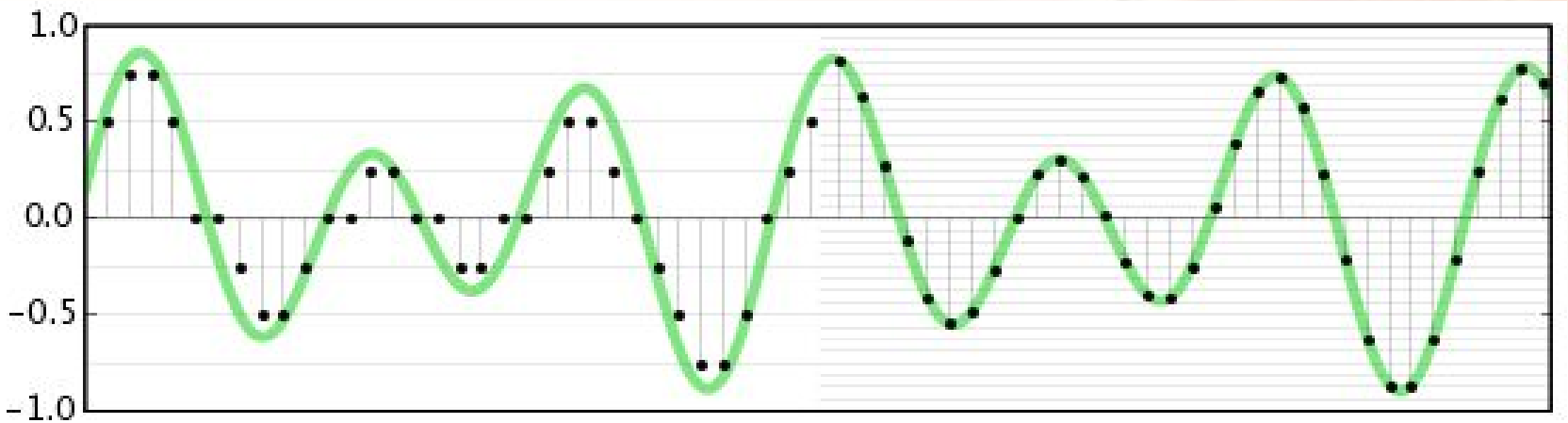


Image adapted from
http://manual.audacityteam.org/man/digital_audio.html

Motivation: Anwendungen

Audio Editor

Ohne?

- Klicken und Abspiele

Mit?

- Bestimmten Eigenschaften können markiert werden
  - Solos, Applaus, Reden, usw.

Sicherheitsdienste

Ohne?

- Hinschauen/hören und warten
- Monotones Arbeit lässt sich gut automatisieren

Mit?

- Gefährliche Geräusche “lernen”

Ton Synthese

Ohne?

- Töne erzeugen und hinhören

Mit?

- Töne erzeugen und vergleichen anhand klar definierten Eigenschaften

Agenda

- Einleitung/Motivation
- **Muscle Fish Verfahren**
- Baum-Verfahren
- Vergleich
- Fazit

Muscle Fish

Generelles Paradigma

1. Genauen Zahlen in einer Ausschnitt
2. Fuzzy search
3. Objektive und subjektive Eigenschaften
4. Musikalischer bzw. sprachlicher Inhalt

Muscle Fish

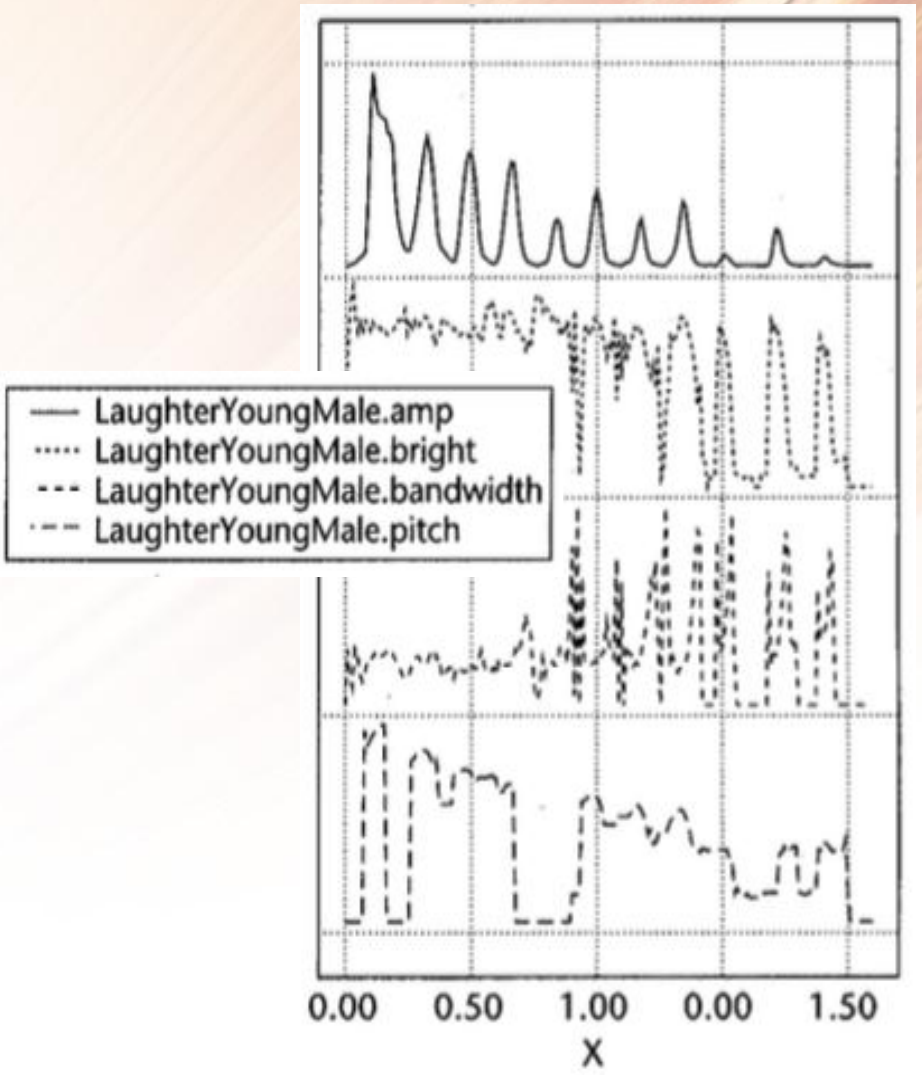
Relevanter Eigenschaften

- Lautheit
  - RMS von Signal
- Tonhöhe
  - Größter gemeinsamer Teiler
- Helligkeit
  - Menge an höheren Frequenzen
- Bandbreite
  - wie viele Frequenzen in der Signal sind
- Harmonizität
  - wie “harmonisch” ein Ton ist

Muscle Fish

Charakterisierung eines Signals

- N-dimensionaler Vektor
  - Mittelwert
  - Varianz
  - Autokorrelation für eine kleine Zeitverschiebung
  - mit der Lautheit gewichtet
- Klassen erstellen
- Töne zuordnen



Adapted from Wold, Erling, et al. "Content-based classification, search, and retrieval of audio."

Muscle Fish

Datenbankeintrag

Analysis feature vector
- Duration
- Pitch [mean, variance, autocorrelation]
- Amplitude [mean, variance, autocorrelation]
- Brightness [mean, variance, autocorrelation]
- Bandwidth [mean, variance, autocorrelation]

Sound file attributes
- File name
- Sample rate
- Sample size
- Sound file format
- Number of channels
- Creation date
- Analysis date

User attributes
- Keywords
- Comments

Klassen erstellen

- Mittelwert und Covarianz Vektoren/Matrizen

$$\mu = (1/M) \sum_j a[j]$$

$$R = (1/M) \sum_j (a[j] - \mu) \cdot (a[j] - \mu)^T$$

Property	Mean	Variance	Autocorrelation
Loudness	−54.4112	221.451	0.938929
Pitch	4.21221	0.151228	0.524042
Brightness	5.78007	0.0817046	0.690073
Bandwidth	0.272099	0.0169697	0.519198

Adapted from Wold, Erling, et al. "Content-based classification, search, and retrieval of audio."

Muscle Fish

Töne zuordnen

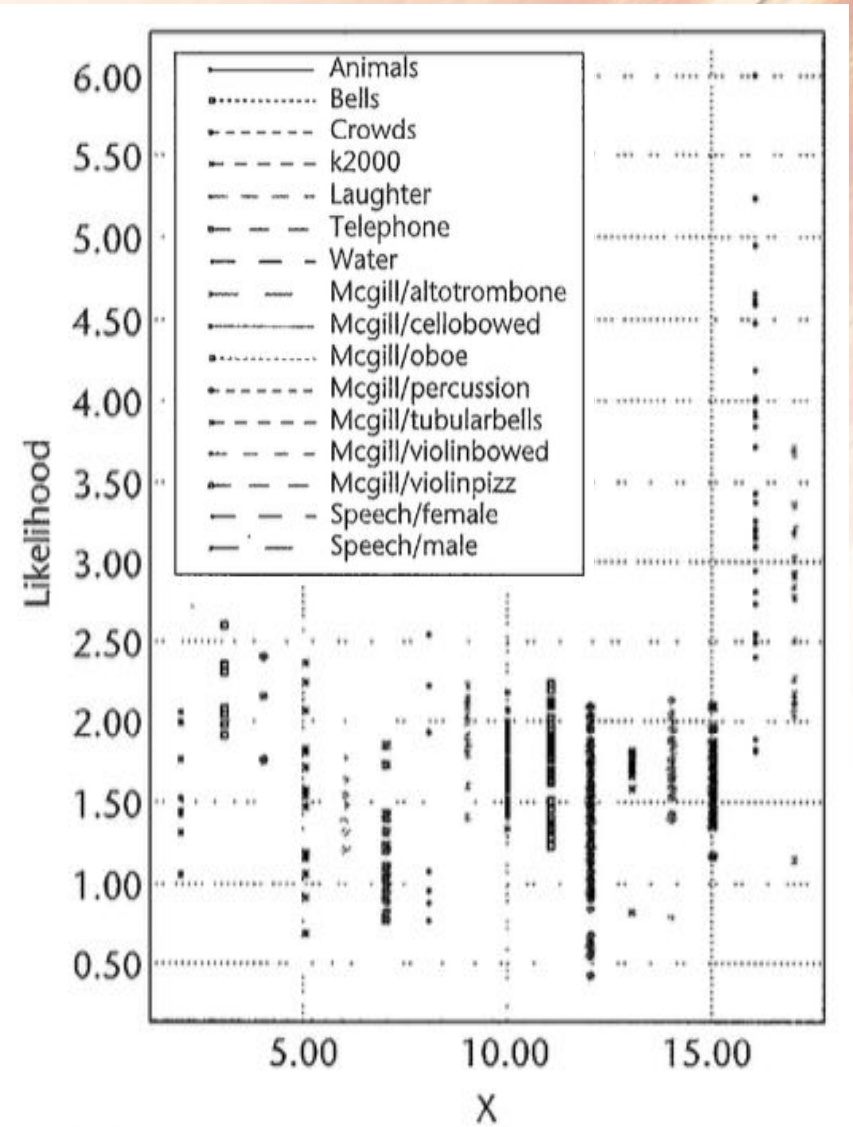
- Gewichteter Euklidischer Abstand

$$D = ((a-\mu)^T R^{-1} (a-\mu))^{1/2}$$

- Likelihood nach der Normalverteilung

$$L = exp(-D^2/2)$$

- “Tear gas” Beispiel



Adapted from Wold, Erling, et al. "Content-based classification, search, and retrieval of audio."

Muscle Fish

- Ob-/Subjektive Eigenschaften
- Statistisches Verfarhen
- Werden durch Vektoren/Abstände beschreiben/verglichen

Agenda

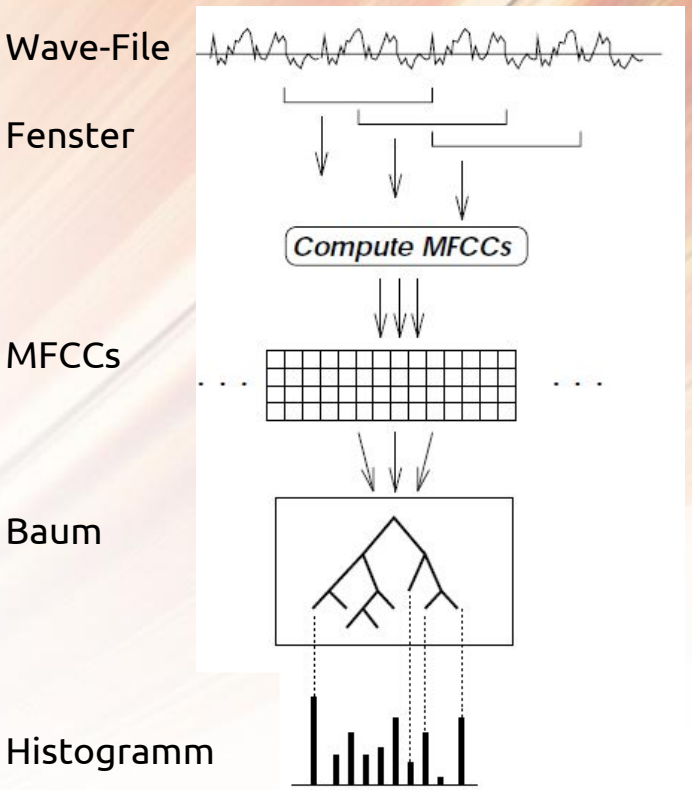
- Einleitung/Motivation
- Muscle Fish Verfahren
- **Baum-Verfahren**
- Vergleich
- Fazit

Baum-Verfahren

Rein datenorientierter Ansatz

Idee: Gegeben eine (ggf. auch unbekannte) Audio-Datei - welche anderen bekannten Audio-Dateien sind ähnlich

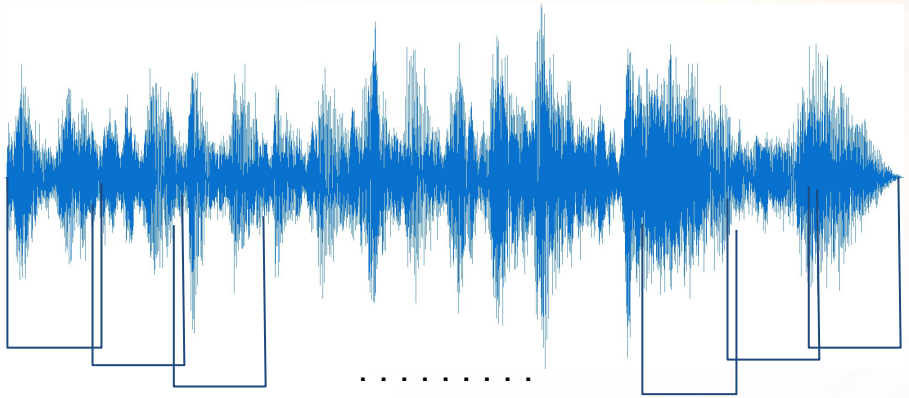
Frage: Wie definiert man Ähnlichkeit bei Audiodateien?

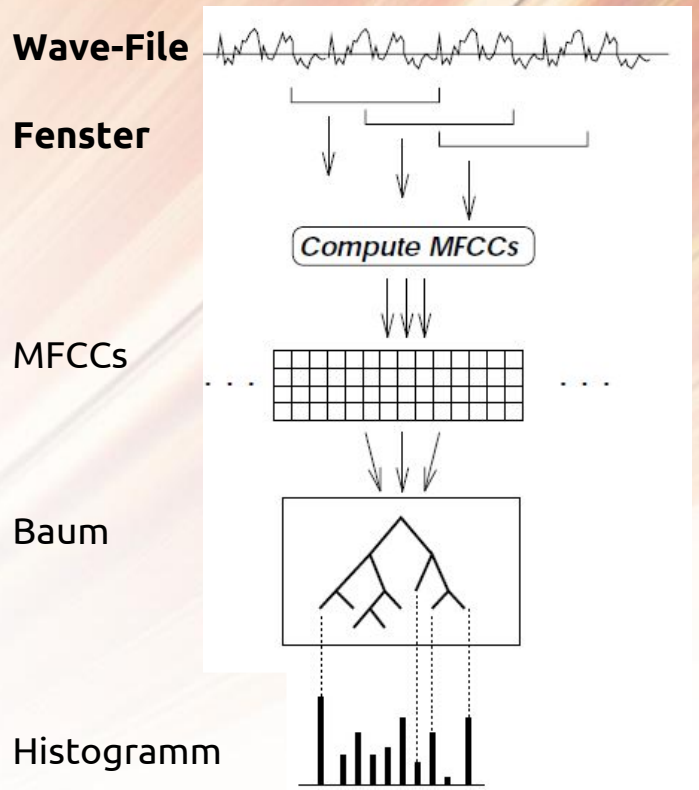


Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

Baum-Verfahren

Waveform Audio File Format (WAVE/WAV)

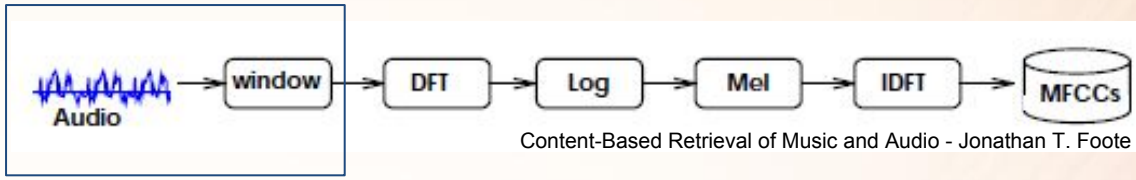




Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

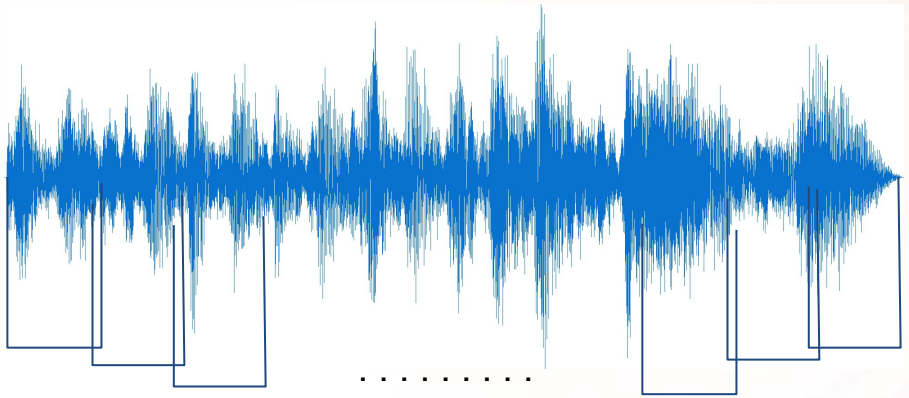
<u>Gängige Parameter:</u>
Fenstergröße: 25ms
Überschneidung: 10ms

Baum-Verfahren



Content-Based Retrieval of Music and Audio - Jonathan T. Foote

Waveform Audio File Format (WAVE/WAV)



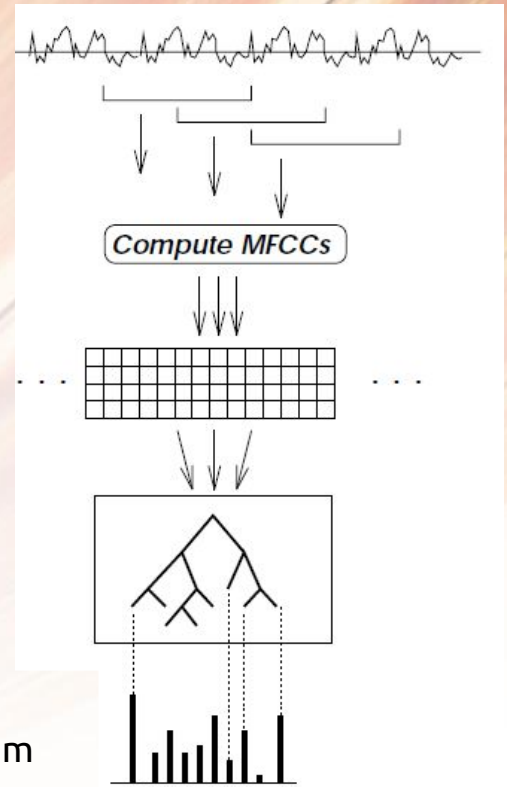
Wave-File

Fenster

MFCCs

Baum

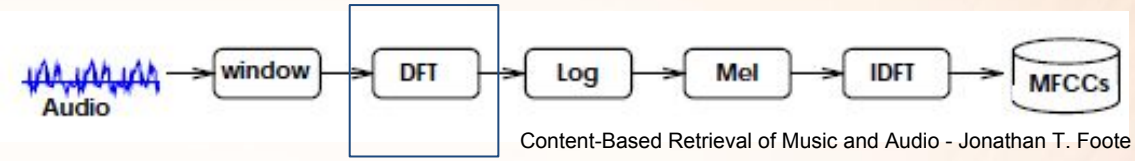
Histogramm



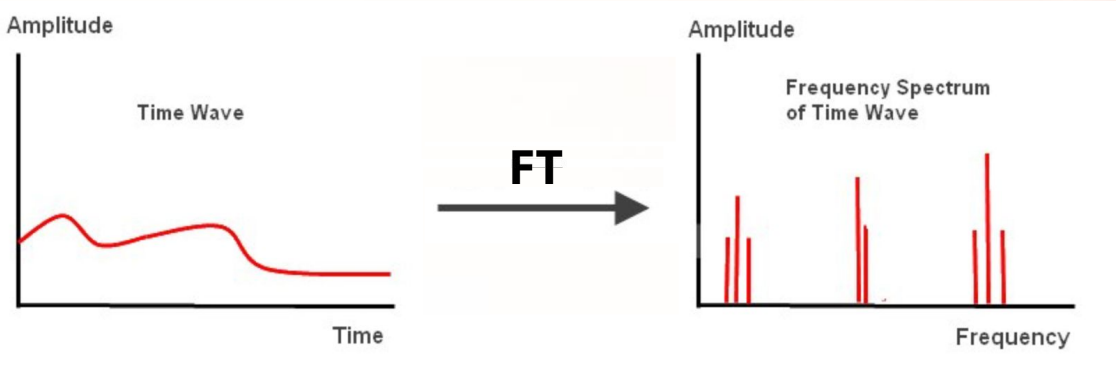
Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

Mel **F**requency **C**epstral **C**oefficients

Baum-Verfahren



Content-Based Retrieval of Music and Audio - Jonathan T. Foote



http://www.alwayslearn.com

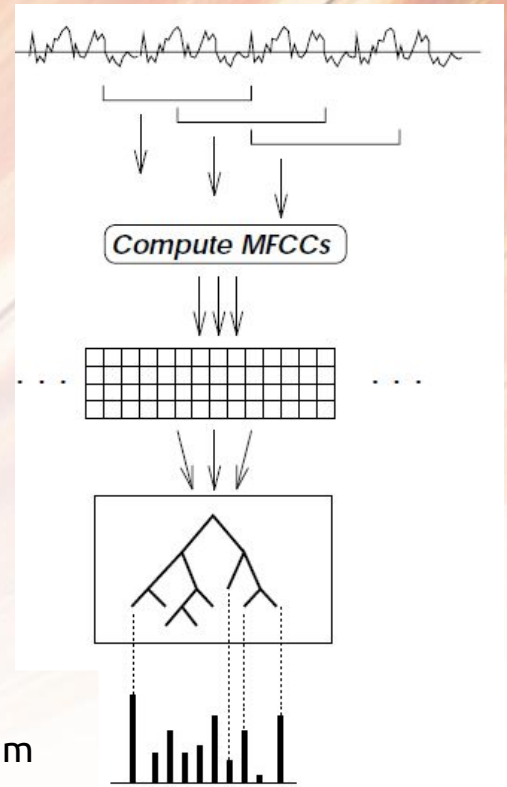
Wave-File

Fenster

MFCCs

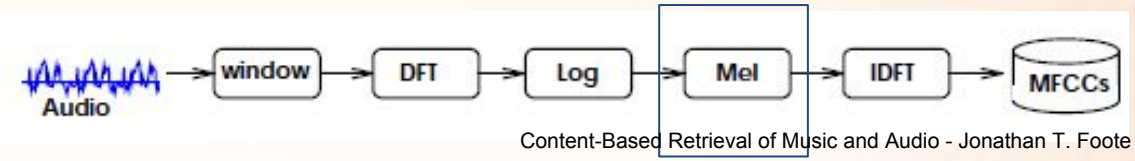
Baum

Histogramm

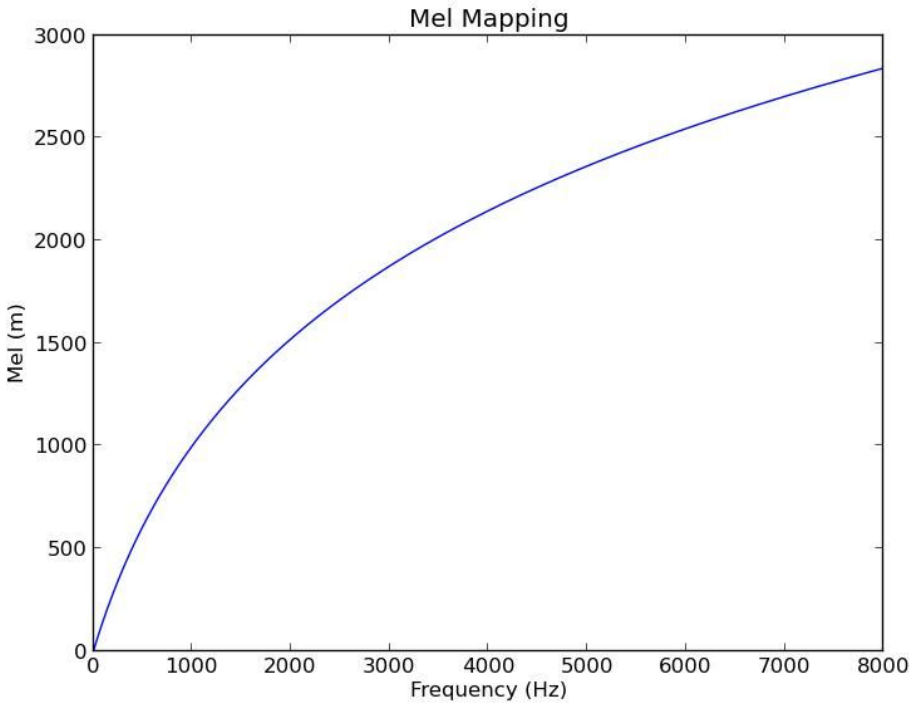


Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

Baum-Verfahren



Content-Based Retrieval of Music and Audio - Jonathan T. Foote



https://www.vocal.com/wp-content/uploads/2013/06/28.png

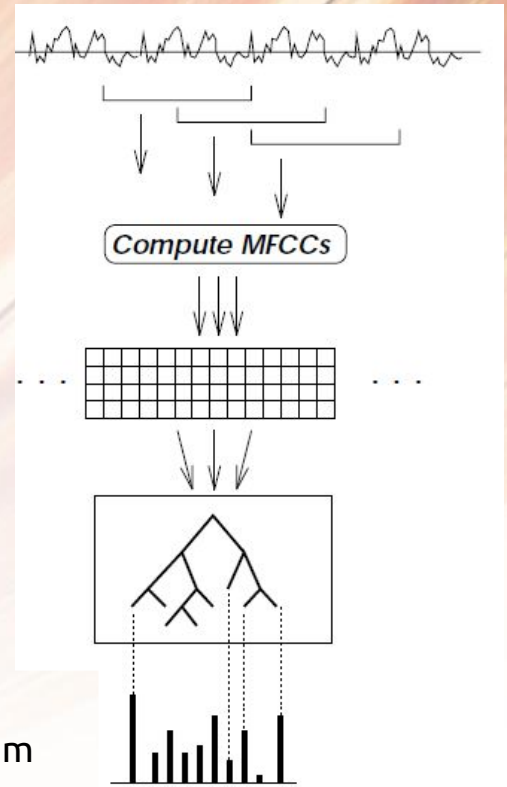
Wave-File

Fenster

MFCCs

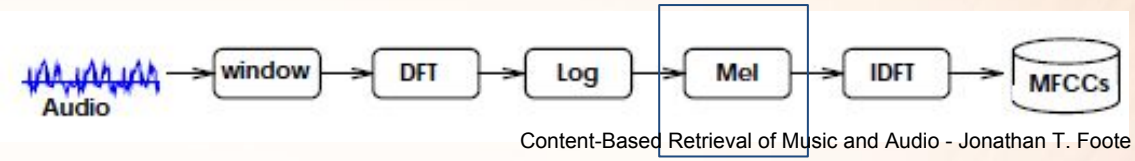
Baum

Histogramm

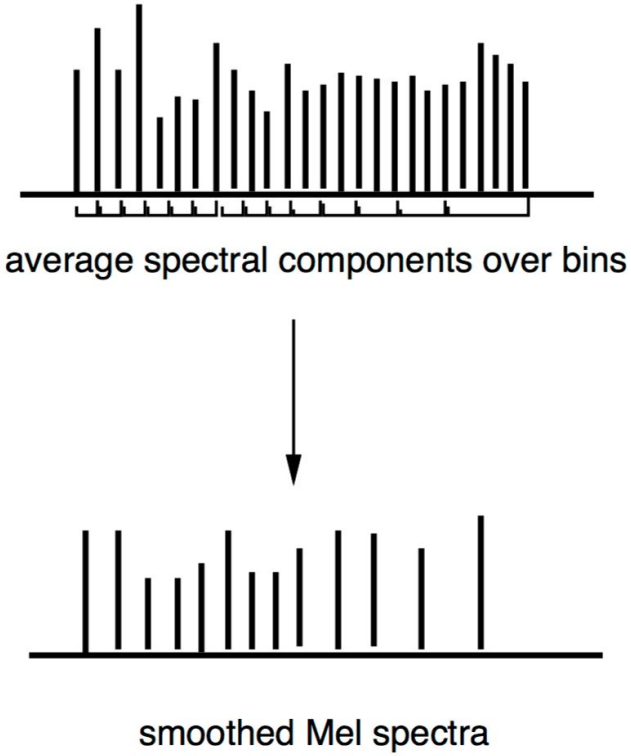


Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

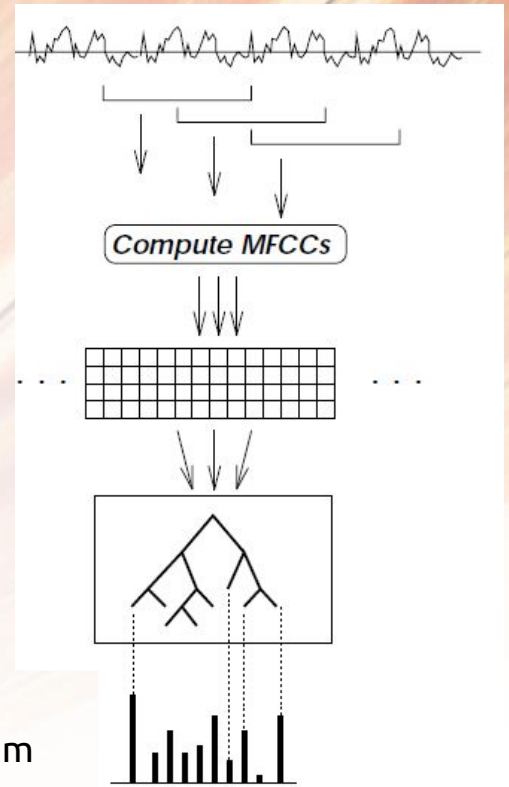
Baum-Verfahren



Content-Based Retrieval of Music and Audio - Jonathan T. Foote

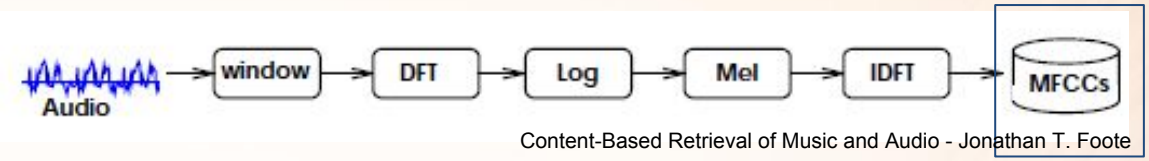


mel frequency cepstral coefficients for music modeling - Beth Logan



Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

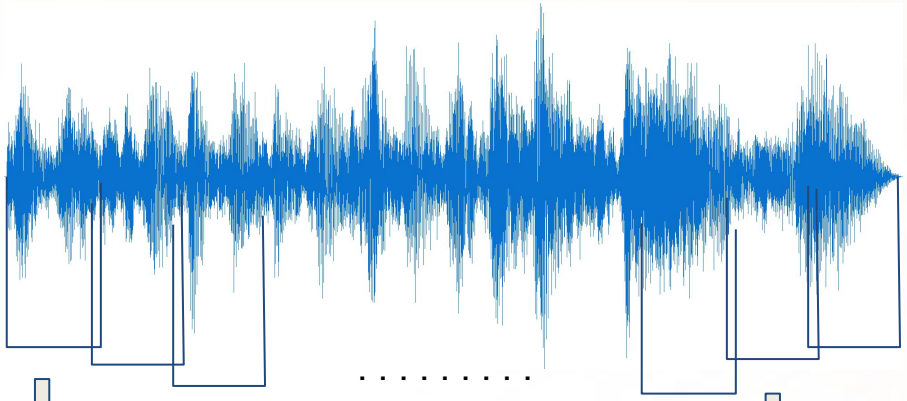
Baum-Verfahren



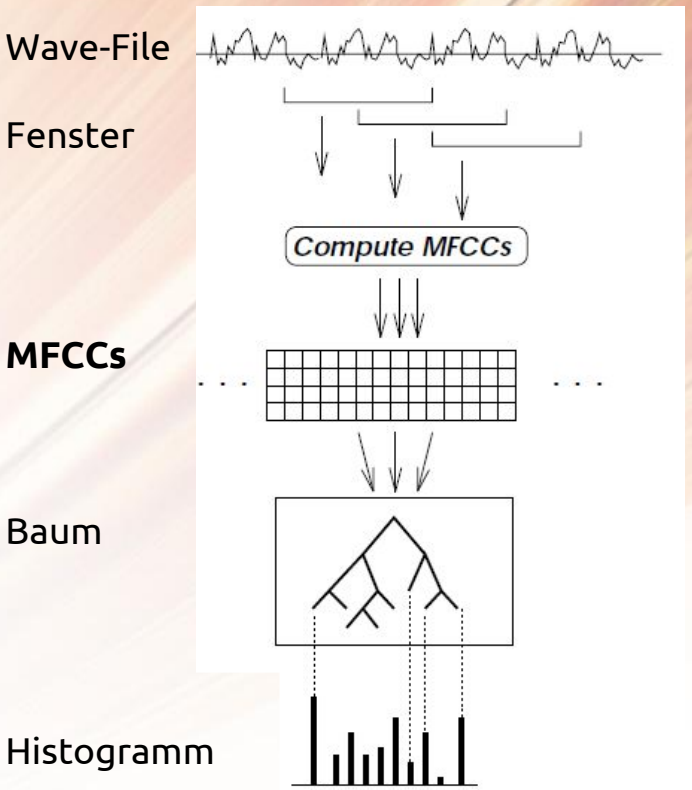
Content-Based Retrieval of Music and Audio - Jonathan T. Foote

Ergebnis:

- Kompakte Darstellung des Frequenzspektrums mit festgelegter Anzahl an Parametern



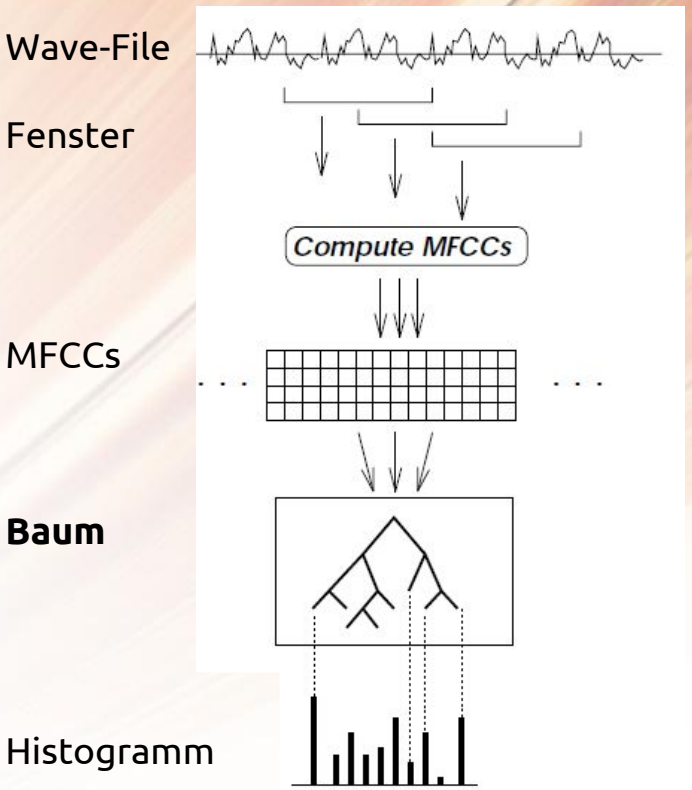
$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix}, \quad \begin{pmatrix} a_{1\,(n-1)} \\ a_{2\,(n-1)} \\ \vdots \\ a_{n\,(n-1)} \end{pmatrix}$$



Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

Baum-Verfahren

- Trainingsdaten klassifiziert
- Supervised Algorithmus
- Ziel: Wahrscheinlichkeit in Blättern maximieren

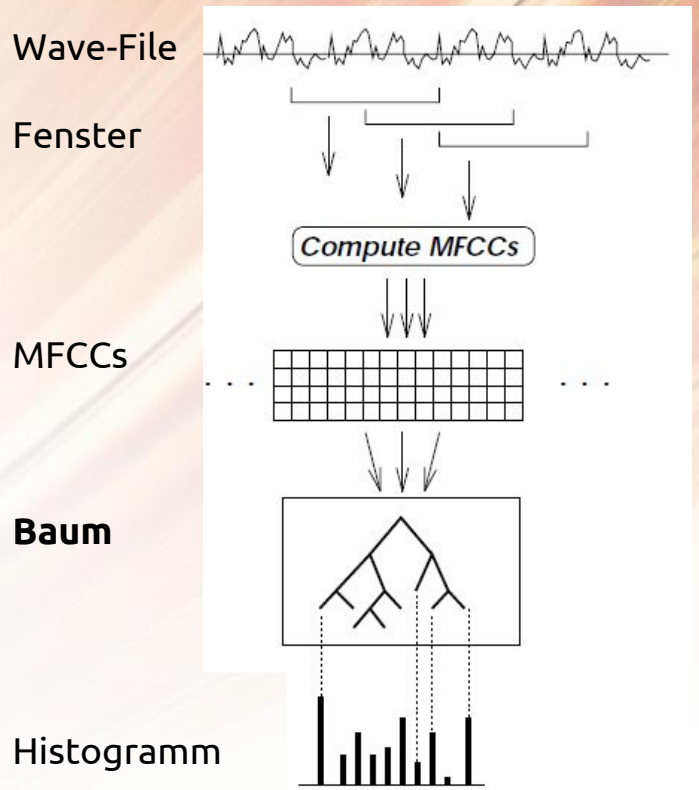


Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

Baum-Verfahren

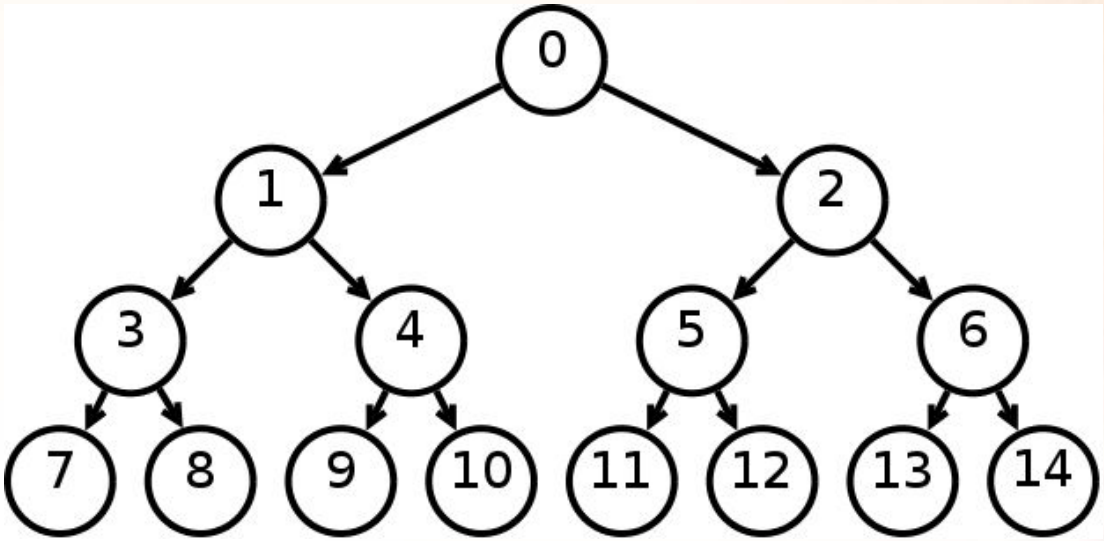
Classification Trees

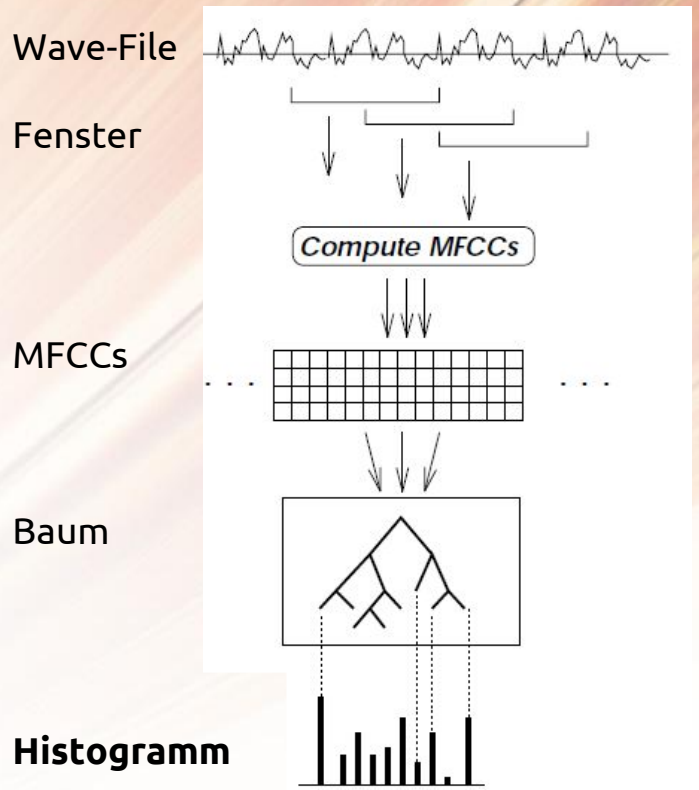
Outlook	Temp (°F)	Humidity (%)	Windy?	Class
Sunny	75	70	True	Play tennis
Sunny	80	90	True	Don't play
Sunny	85	85	False	Don't play
Sunny	72	95	False	Don't play
Sunny	69	70	False	Play tennis
Overcast	72	90	True	Play tennis
Overcast	83	78	False	Play tennis
Overcast	64	65	True	Play tennis
Overcast	81	75	False	Play tennis
Rain	71	80	True	Don't play
Rain	65	70	True	Don't play
Rain	75	80	False	Play tennis
Rain	68	80	False	Play tennis
Rain	70	96	False	Play tennis



Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

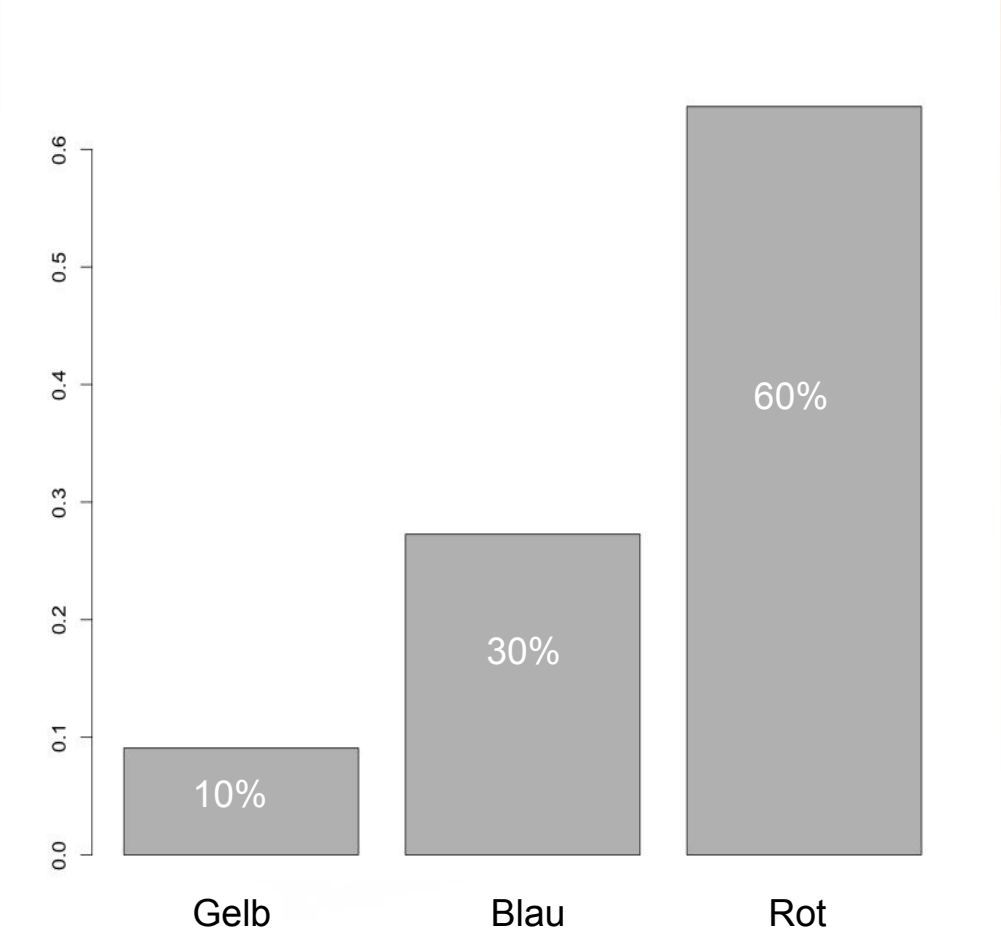
Baum-Verfahren

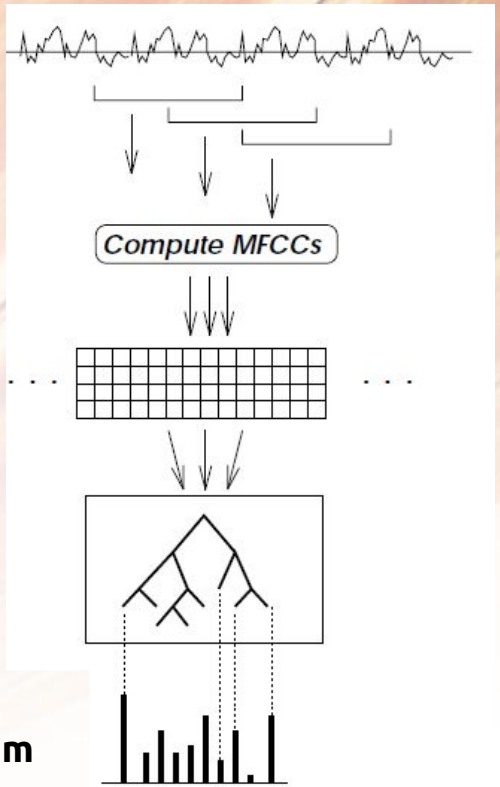




Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

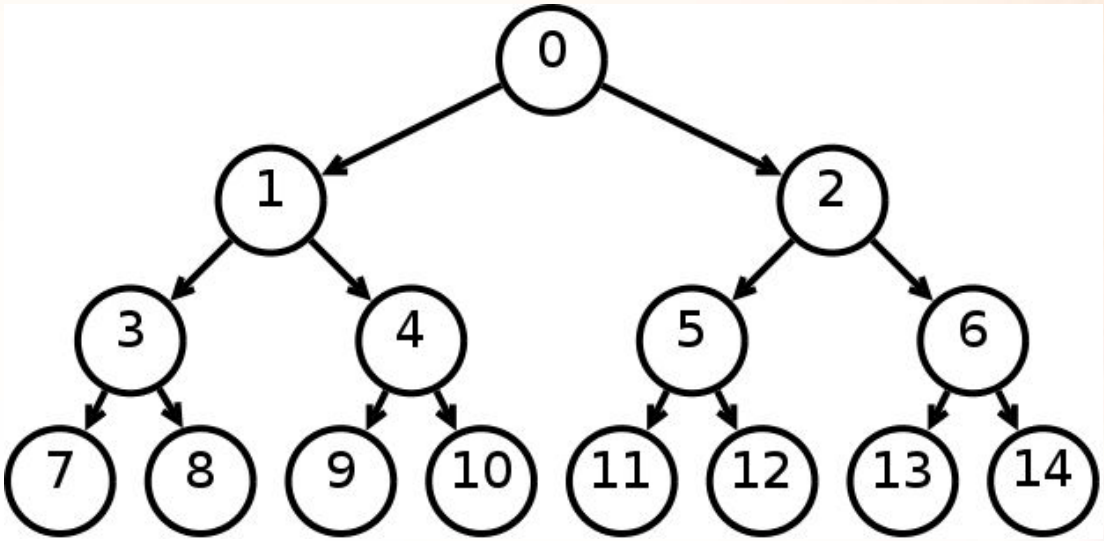
Baum-Verfahren



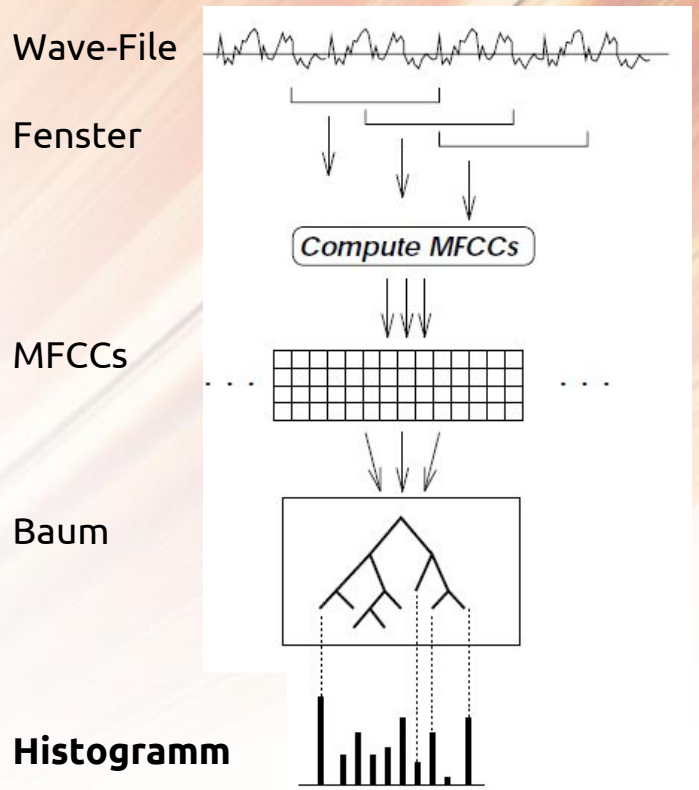


Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

Baum-Verfahren

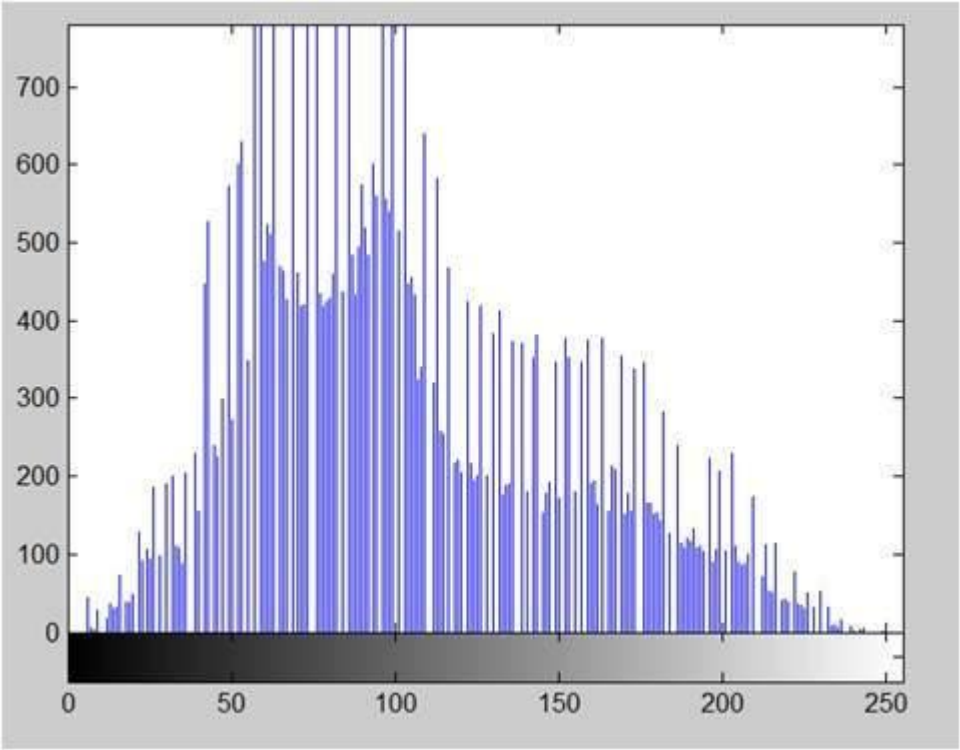


https://hbfs.files.wordpress.com/2009/04/diagram1-2.png

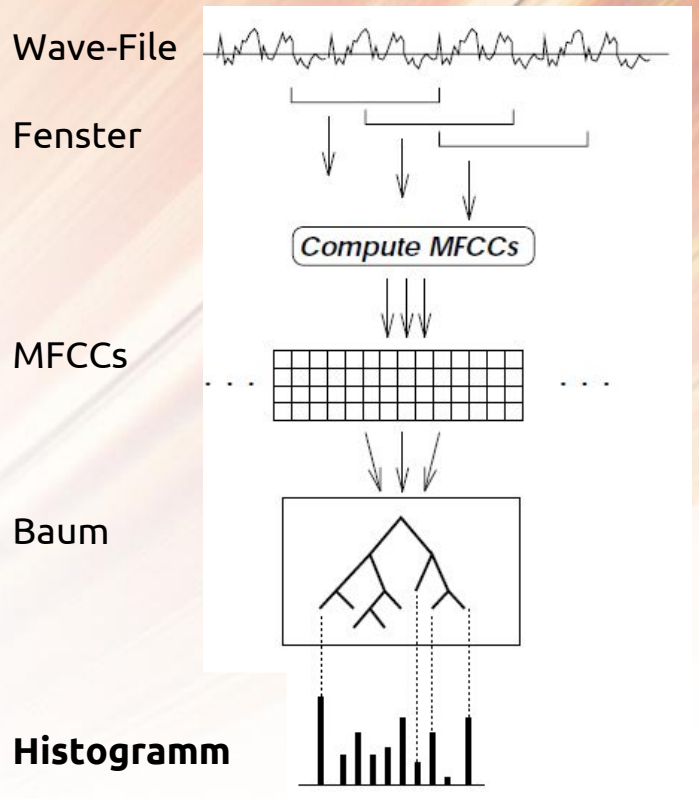


Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

Baum-Verfahren



https://www.tutorialspoint.com/dip/images/histogram3.jpg



Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

Baum-Verfahren

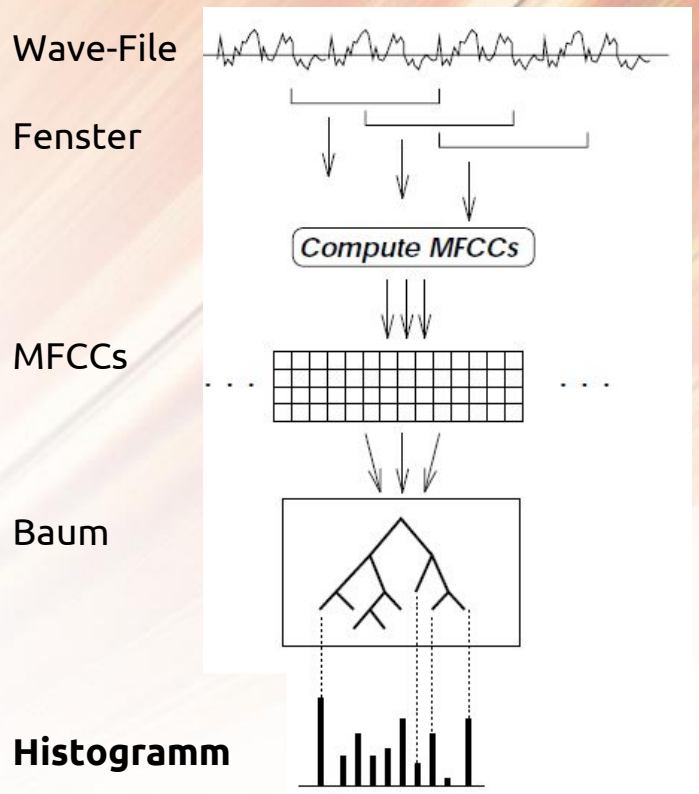
Eingangsfrage:
Wann sind zwei Audiosignale ähnlich?

Antwort:
Wenn ihre Histogramme ähnlich sind!

Neue Frage:
Wie wird die Ähnlichkeit von Histogrammen definiert?

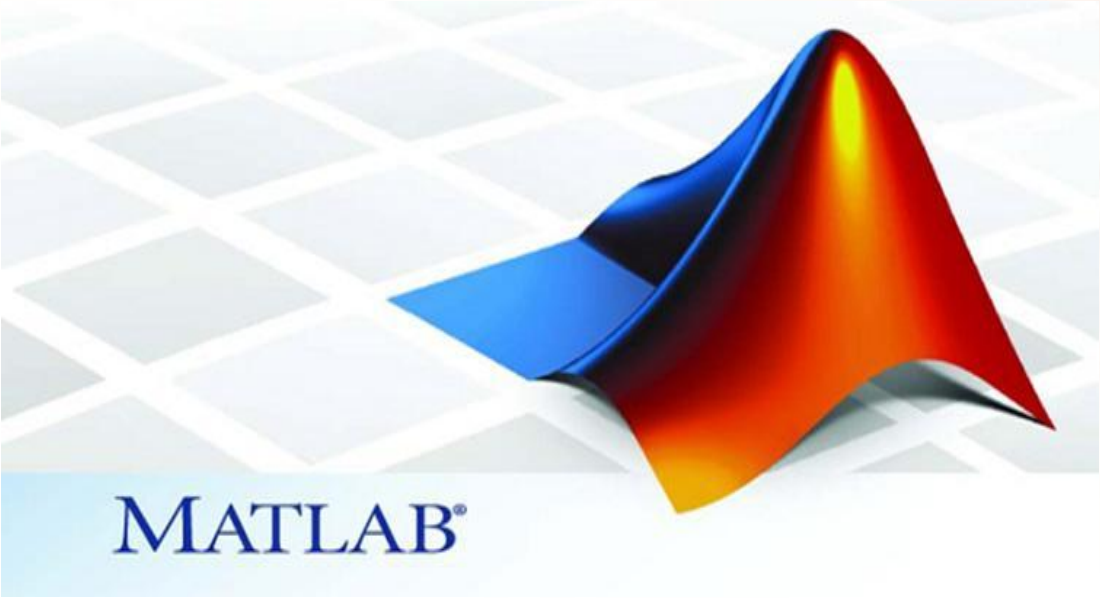
Antwort:
Euklidischer Abstand!

$$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$



Adapted from Content-Based Retrieval of Music and Audio - Jonathan T. Foote

Demo



http://www.citlprojects.com

Agenda

- Einleitung/Motivation
- Muscle Fish Verfahren
- Baum-Verfahren
- **Vergleich**
- Fazit

Vergleich

- Test mit 409 Audiodateien
- “Supervektoren” mit 65 Features (5 * 13)
- Supervised: 41 Klassen (Oboe, ..)

Distance	Q Tree (D_C) “unsupervised”	Q Tree (D_C) supervised	Muscle Fish (no DPL)	Muscle Fish (+ DPL)
Laughter (M)	0.68	0.82	1.00	1.00
Oboe	0.11	0.43	0.69	0.94
Agogo	1.00	1.00	0.53	0.58
Speech (F)	0.77	0.87	0.69	0.94
Touchtone	0.61	1.00	0.44	0.73
Rain/thunder	0.22	0.35	0.30	0.42
Mean AP	0.580	0.772	0.608	0.768

Content-Based Retrieval of Music and Audio = Jonathan T. Foote

Agenda

- Motivation/Einleitung
- Muscle Fish Verfahren
- Baum-Verfahren
- Vergleich
- **Fazit**

Fazit

- Es funktioniert.
- Ergebnisse / Modelle stark abhängig von den verwendeten Daten.
- Viele Parameter können Ergebnis beeinflussen.
- Unsupervised Algorithmen schon recht Präzise.

Quellen

- Logan, Beth. "Mel Frequency Cepstral Coefficients for Music Modeling." *ISMIR*. 2000.
- Wold, Erling, et al. "Content-based classification, search, and retrieval of audio." *IEEE multimedia* 3.3 (1996): 27-36.
- Foote, Jonathan T. "Content-based retrieval of music and audio." *Voice, Video, and Data Communications*. International Society for Optics and Photonics, 1997.
- Sigurdsson, Sigurdur, Kaare Brandt Petersen, and Tue Lehn-Schiøler. "Mel frequency cepstral coefficients: An evaluation of robustness of mp3 encoded music." *Seventh International Conference on Music Information Retrieval (ISMIR)*. 2006.
- Muda, Lindasalwa, Mumtaj Begam, and I. Elamvazuthi. "Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques." *arXiv preprint arXiv:1003.4083* (2010).
- Markus Schedl, Emilia Gomez, Julian Urbano. “Music Information Retrieval: Recent Developments and Applications”. 2014