

The background of the slide features a soft, warm-toned gradient with faint, stylized musical notes and symbols scattered across it. The notes are in shades of yellow, orange, and pink, creating a musical and artistic atmosphere.

Lea Tschiersch

Kristin Knorr

08.12.2016

# Recognition of Coversongs

# Gliederung

- i. Einleitung
- ii. Lösungsansätze
- iii. LSH
- iv. Fazit

# Gliederung

- Einleitung
  - Motivation
  - Verschiedene Covertypen
  - Involvierte musikalische Aspekte
  - Wissenschaftliche Anwendungen

# Motivation

- Informationsgewinnung über Messbarkeit von Ähnlichkeit von Musik
- Wichtiger Aspekt für Musik- und Lizenzrechte
- Organisation von Musikbibliotheken

# Covertypen

- **Remaster:** überarbeitete Veröffentlichung
- **Instrumental:** ohne Gesang
- **Live performance:** Konzertmitschnitt
- **Acoustic:** keine elektrische Verstärkung
- **Demo:** erster Entwurf



# Covertypen

- **Duett:** Neuaufnahme mit zwei Sängern
- **Medley:** Aneinanderreihung verschiedener Songausschnitte
- **Remix:** elektronisch überarbeitetes Stück
- **Quotation:** kurzes musikalisches Zitat

# Musikalische Aspekte

- **Timbre:** generelle Änderung der Klangfarbe
- **Tempo:** Geschwindigkeitsänderung
- **Timing:** rhythmische Änderungen
- **Structure:** strukturelle Änderung

# Musikalische Aspekte

- **Key:** Transposition
- **Harmonization:** Änderung der Harmonik
- **Lyrics:** Änderung des Textes
- **Noise:** Hintergrundgeräusche



# Übersicht Covertypen

**Table 1** Musical changes that can be observed in different cover song categories. Stars indicate that the change is possible, but not necessary.

	Timbre	Tempo	Timing	Structure	Key	Harm.	Lyrics	Noise
Remaster	★							
Instrumental	★						★	★
Live	★	★	★					★
Acoustic	★	★	★		★	★		★
Demo	★	★	★	★	★	★	★	★
Medley	★	★	★	★	★			★
Remix	★	★	★	★	★	★	★	★
Quotation	★			★				★

# Beispiele



Original



Coversongs

# Wissenschaftliche Anwendungen

- Query-by-humming systems
- Content-based music retrieval
- Genre classification
- Audio fingerprinting

# Was bisher geschah...

- Was sind Coversongs und welche Arten gibt es?
- Warum interessiert man sich dafür?

# Was nun?

i. Einleitung

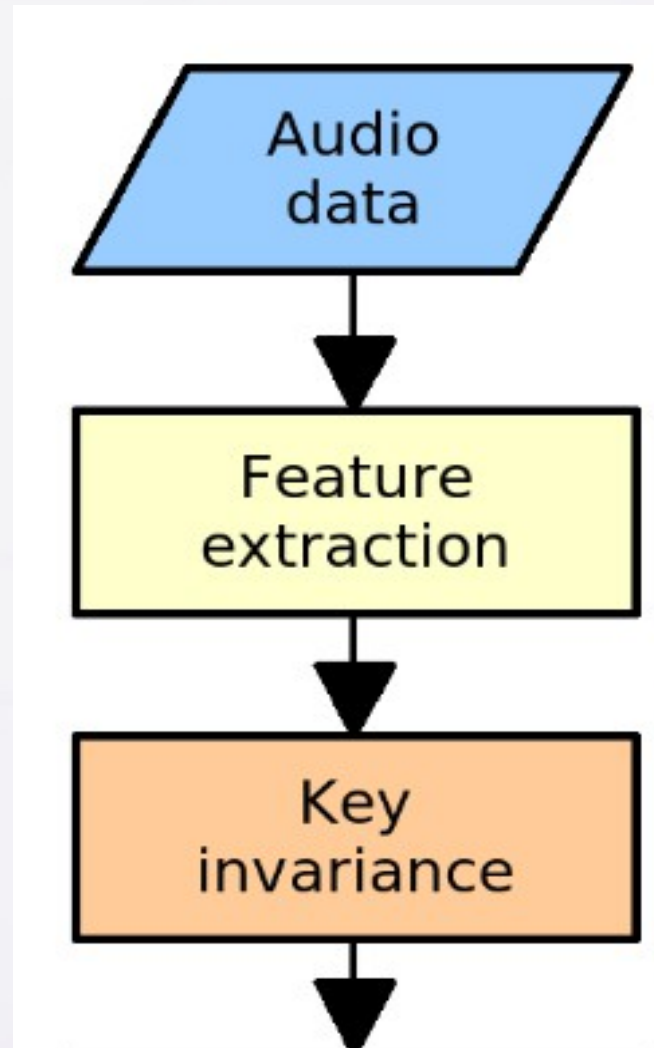
**ii. Lösungsansätze**

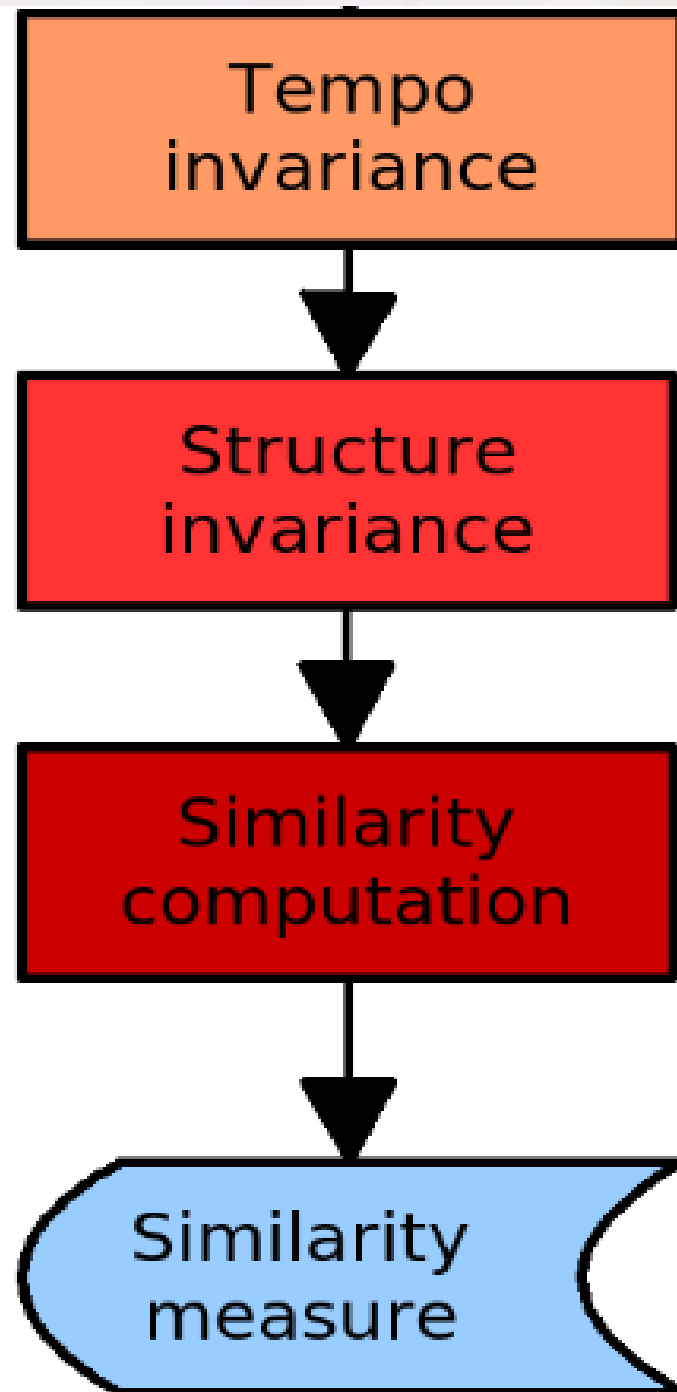
iii. LSH

iv. Fazit



# Ansätze zur Coversong Recognition





# Feature extraction

- Genereller Ausgangspunkt:  
**Versionen des selben Stückes  
enthalten die gleiche Melodie  
und harmonische Entwicklung**

☐ Extraktion der Melodie

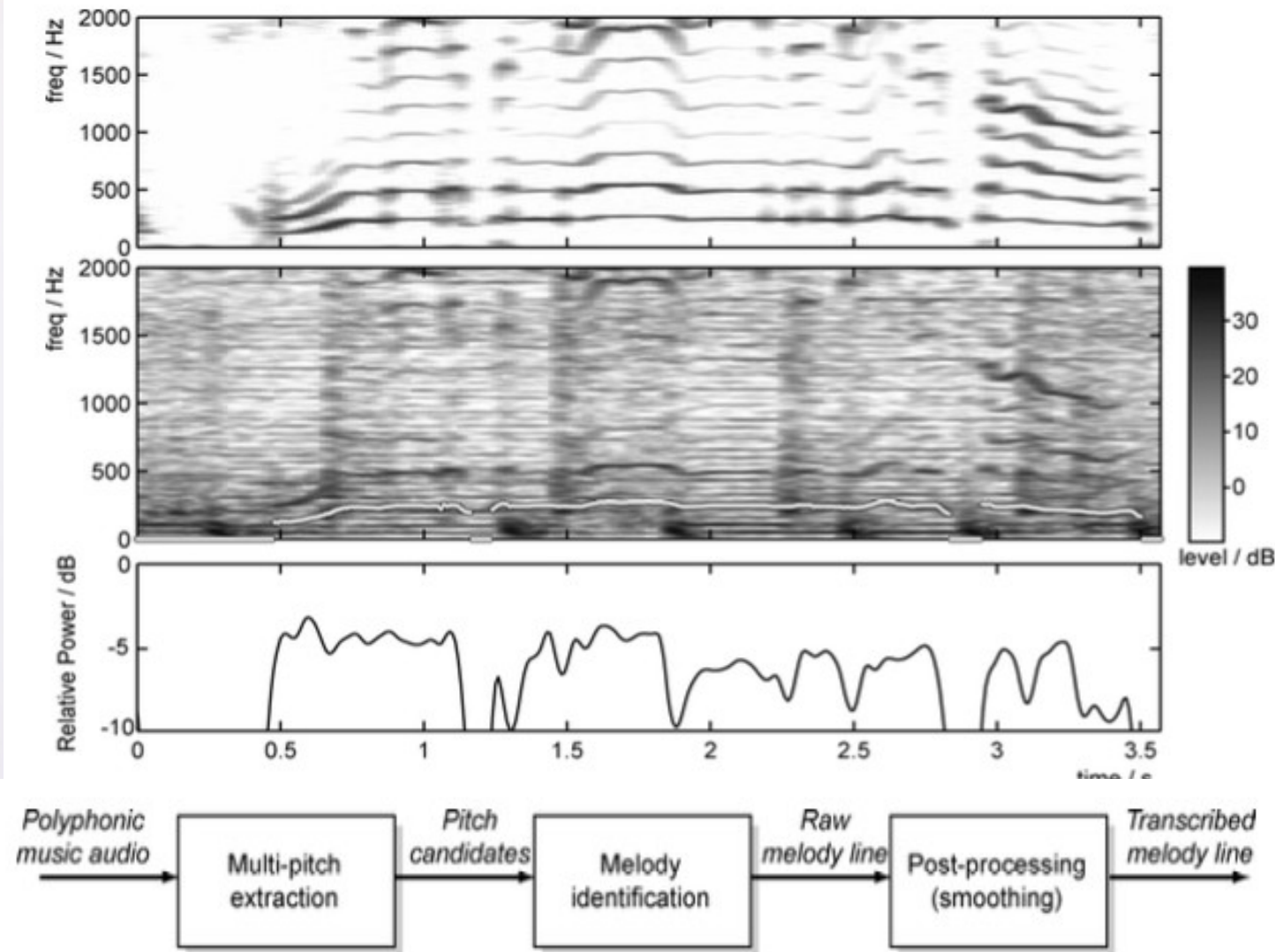


# Extraktion der Melodie

- Entfernen von Pausen und Störgeräuschen
- Ansetzen einer Grenze an z.B. Lautstärke

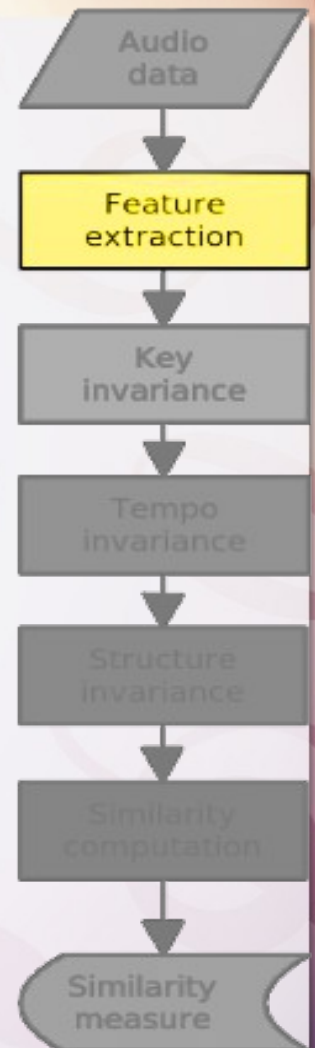
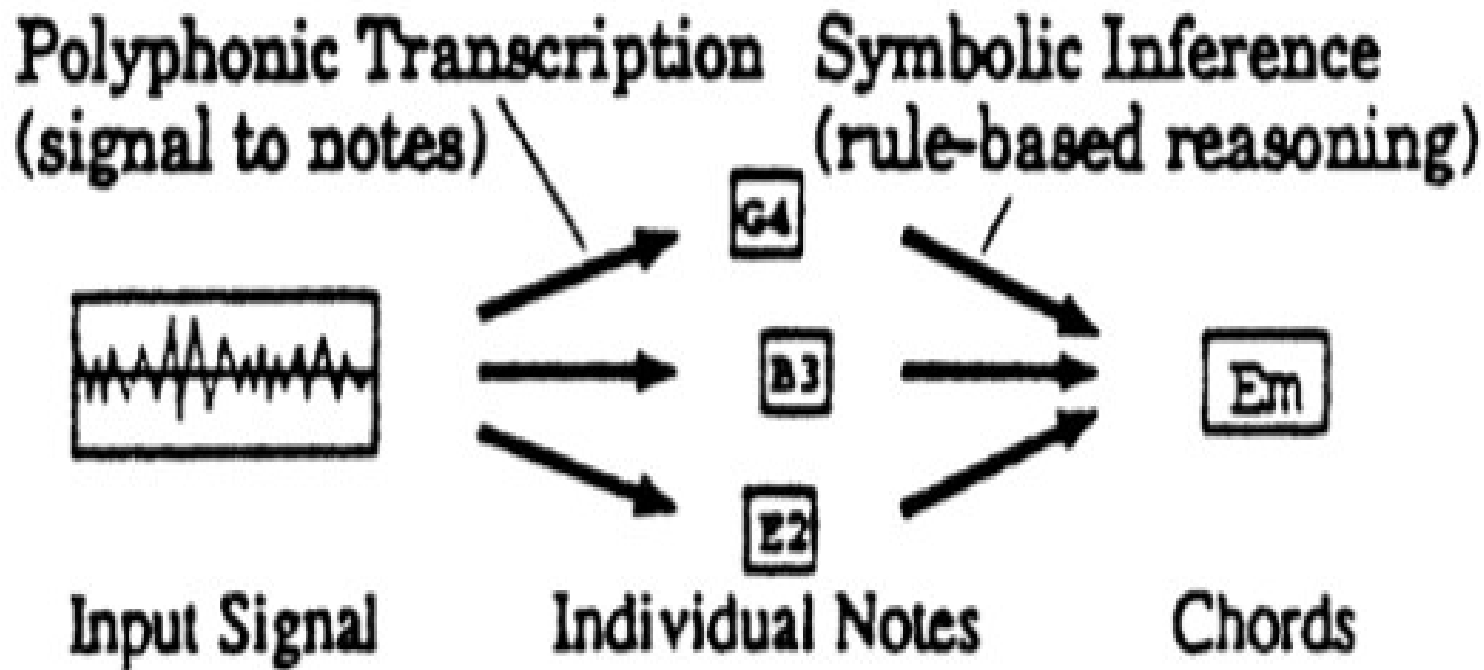


# Extraktion der Melodie





# Akkord-Sequenzen



# PCP – Pitch Class Profiles

- Zuverlässigere, direktere Darstellung als Melodie-Erkennung
- Aufteilung im 100ms Frames
- Bildung eines Vektors, der die Verteilung auf die Tonklassen (Pitch Classes) beschreibt
  - ☐ 12-dimensionaler Vektor



# PCP – Pitch Class Profiles

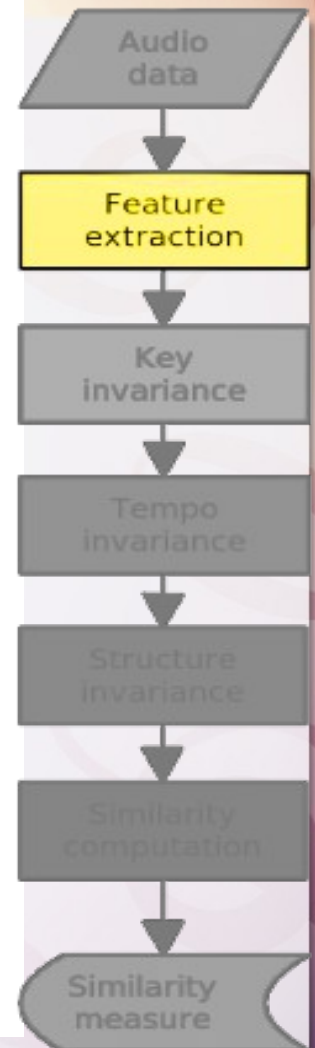
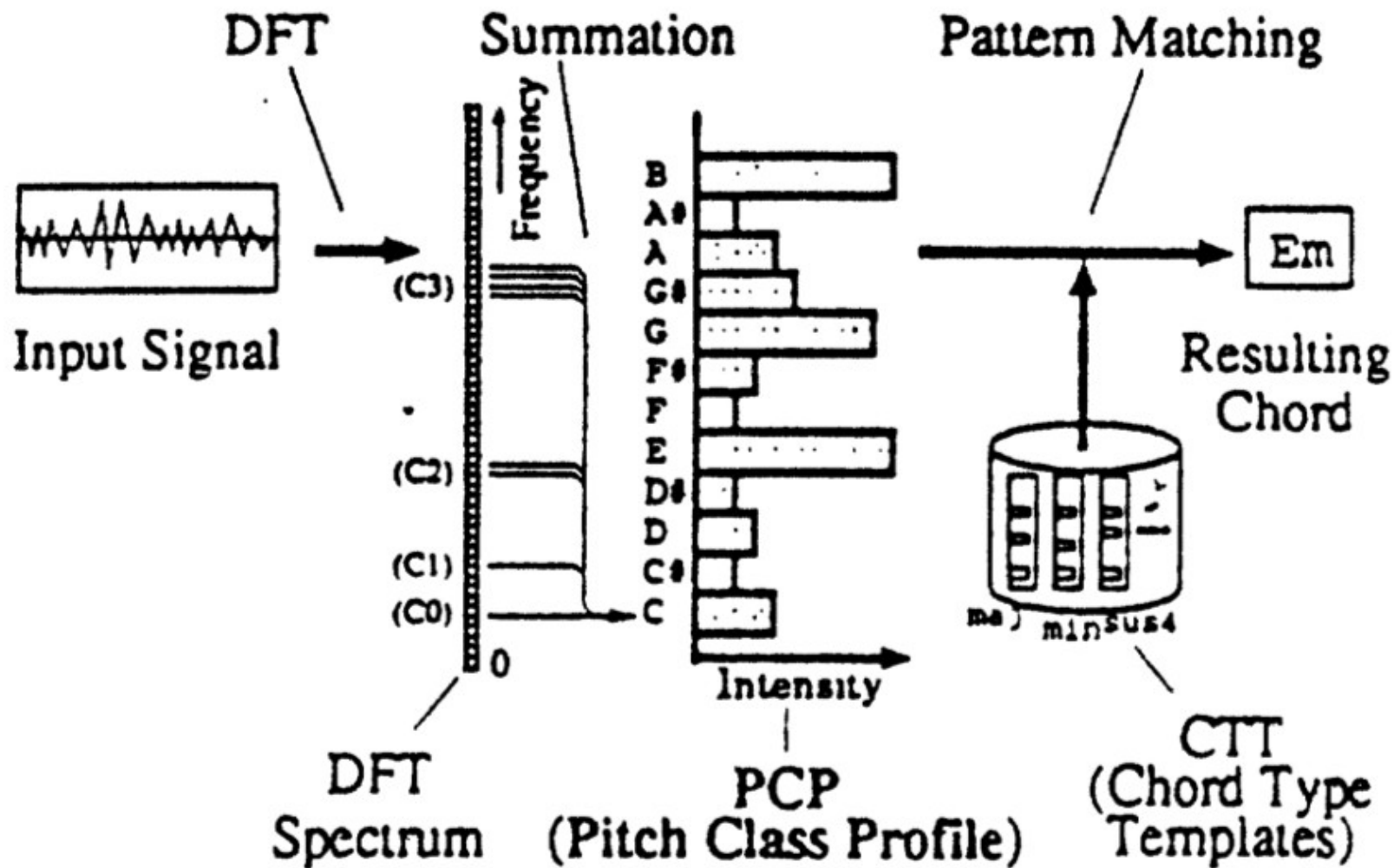
$$X(k) = \sum_{n=0}^{N-1} e^{(-2\pi i k n / N)} \cdot x(n)$$

$$PCP(p) = \sum_l s.t. M(l) = p \|X(l)\|^2$$

$$M(l) = \begin{cases} \text{if } 0, \text{ then } -1 \\ \text{else, round} \left( 12 \log_2 \left( \left( \frac{f_s \cdot 1}{N} \right) / f_{ref} \right) \right) \bmod 12 \end{cases}$$



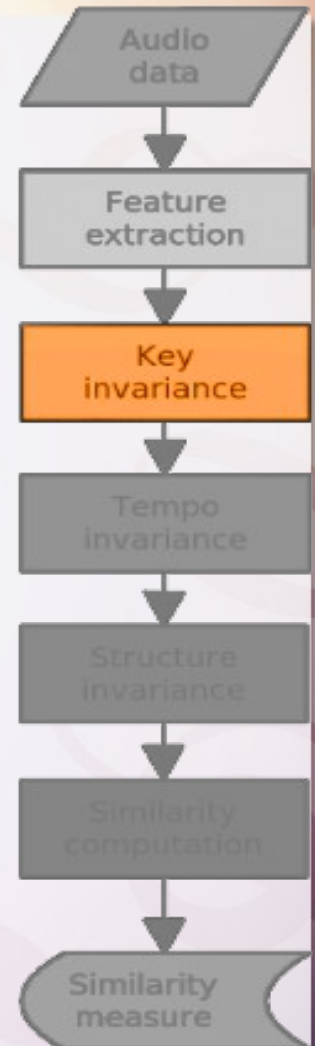
# Vom Signal zum Akkord





# Key Invariance

- Einfachster Weg: Transposition zu allen möglichen Tonarten
- Schneller: Algorithmus zur Berechnung der wahrscheinlichsten Transpositionen mehrerer Songs
- 2D power spectrum
- 2D Autokorrelationsfunktion





# 2D power spectrum

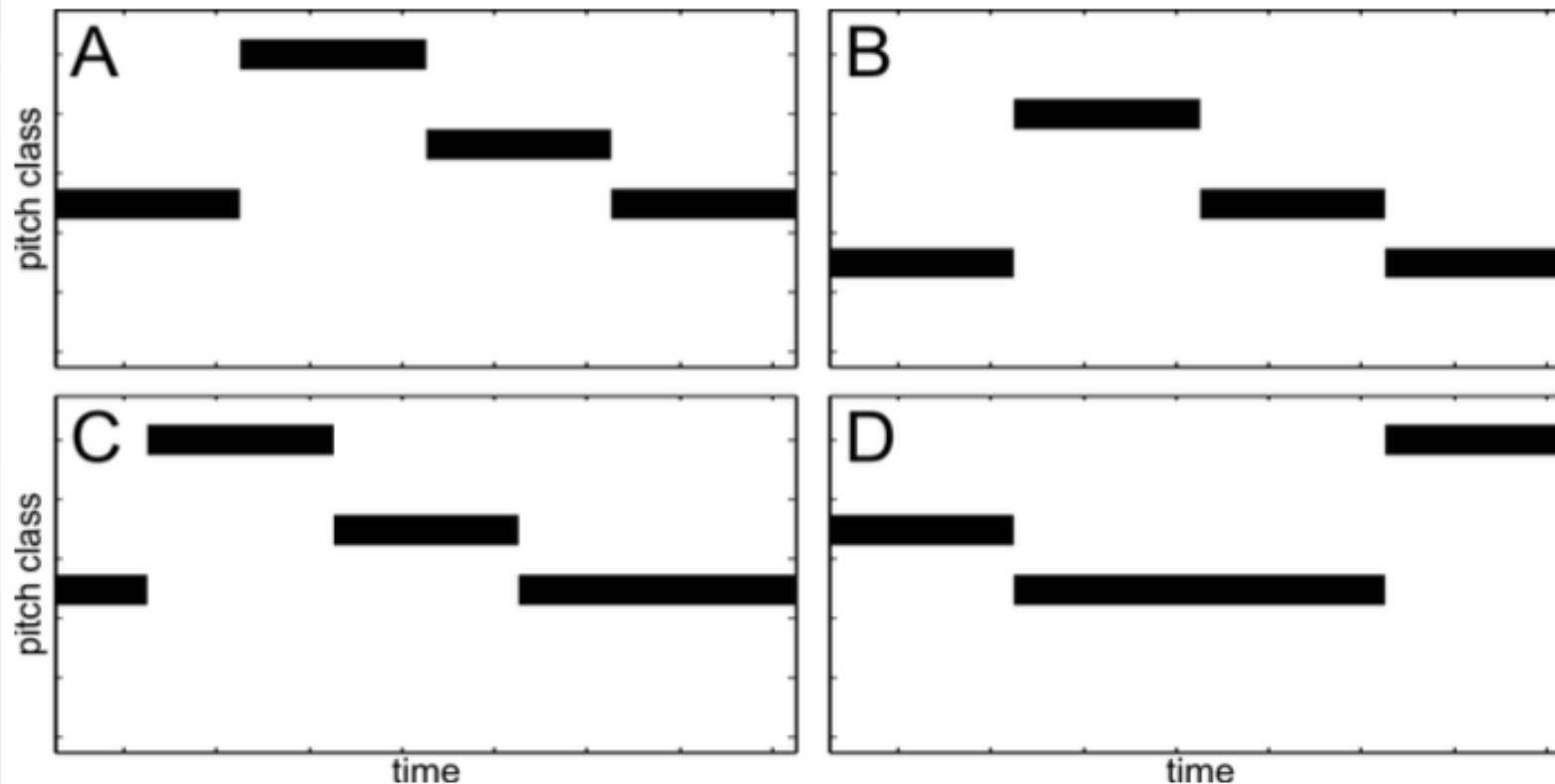
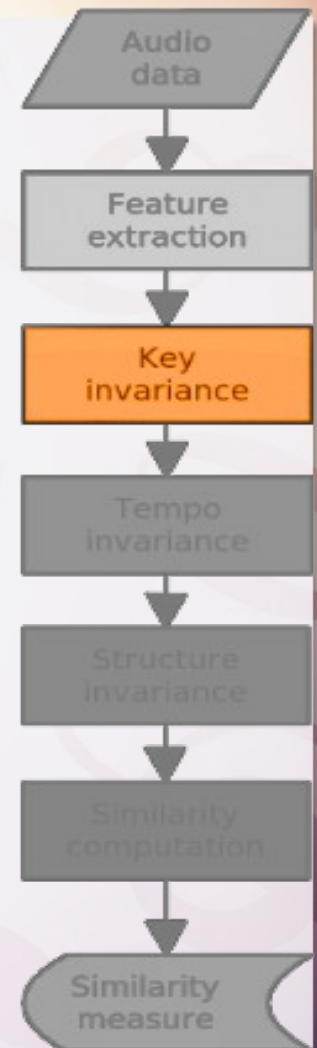


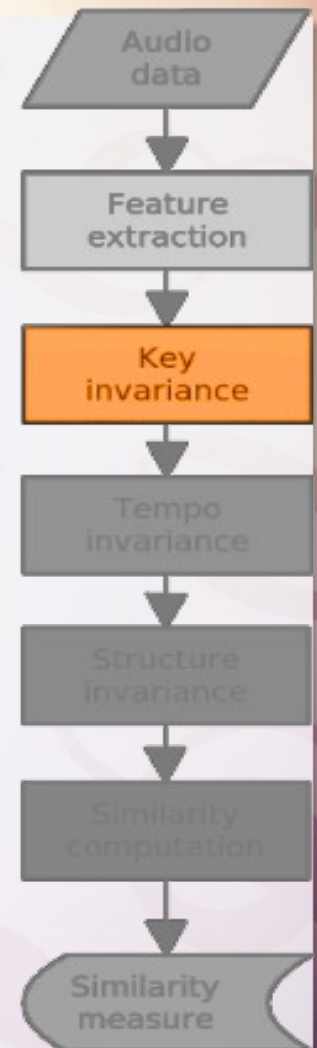
Fig. 3. Shifted melodic fragments with equal power spectra. (A) original, (B) transposed by two semitones, (C) time shifted by 1 beat, and (D) time shifted by 8 beats resulting in a musically different fragment.



# 2D Autokorrelationsfunktion

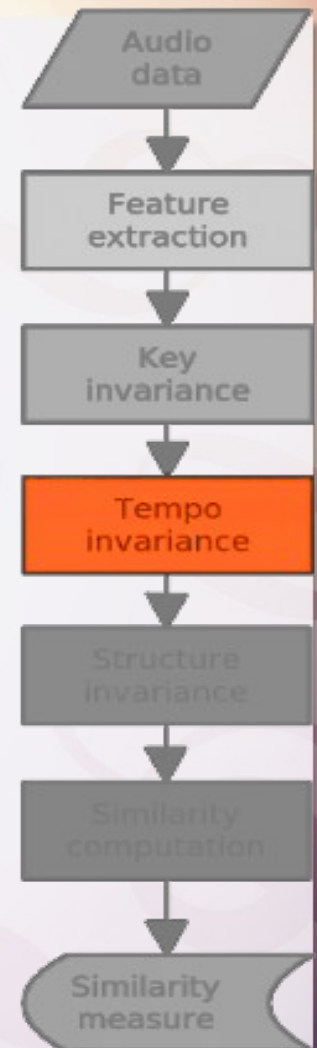
$$R_1(m, k) = \sum_{s=0}^{11} \sum_{n=0}^N c_s(n) c_{(s+m \bmod 12)}(n+k)$$

$$R_2(m, l) = \sum_{k=0}^N R_1(m, k) B_l(k), l \in [0, 15]$$



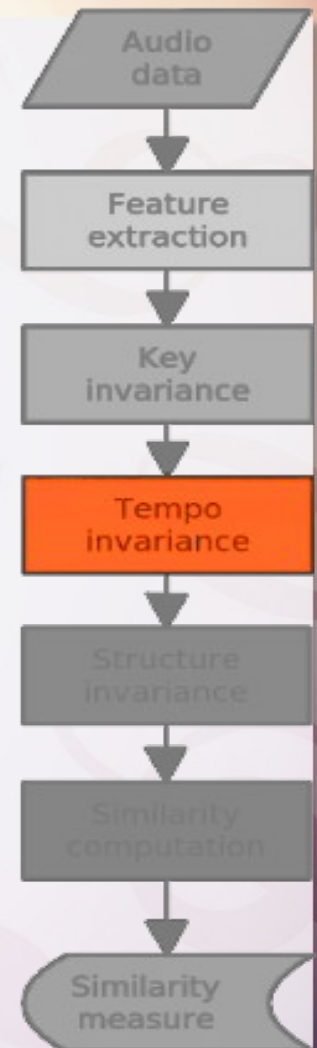
# Tempo invariance

- Schätzung des Tempos durch den Takt
- Zeitliche Kompression/Expansion
- 2D autocorrelation function
- Sequence alignment/similarity algorithm



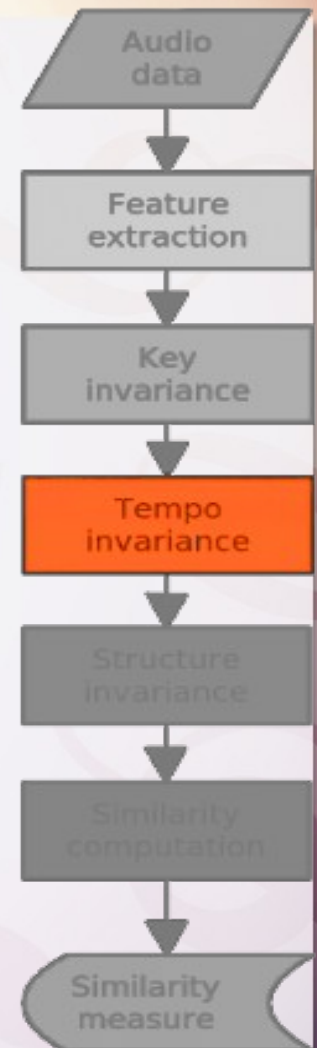
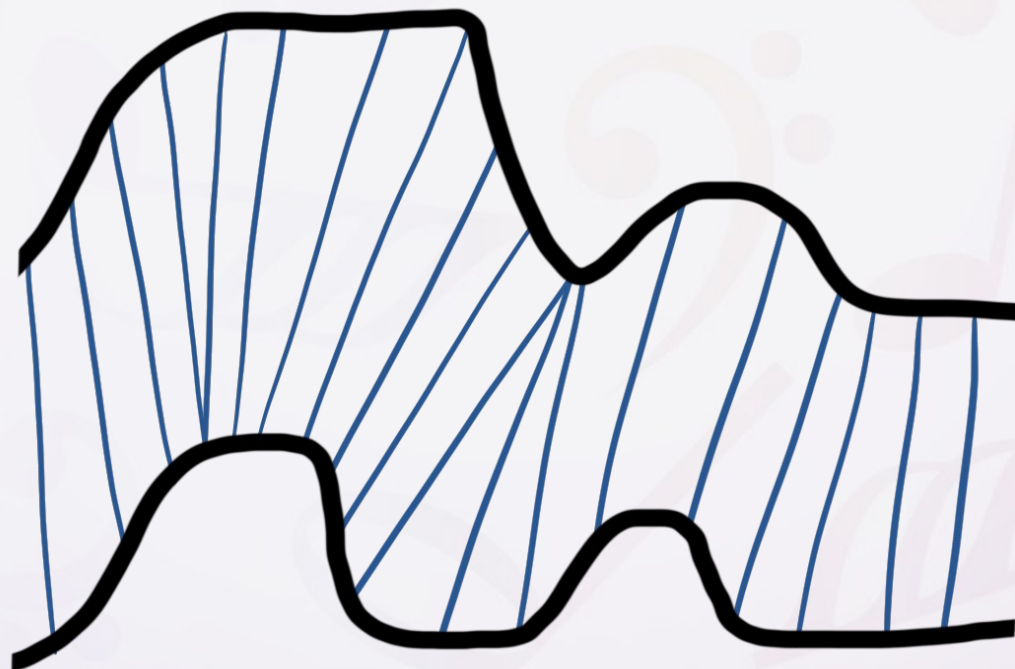
# Dynamic Time Warping

- Algorithmus zum Vergleich von zwei unterschiedlich schnellen Sequenzen



# Dynamic Time Warping

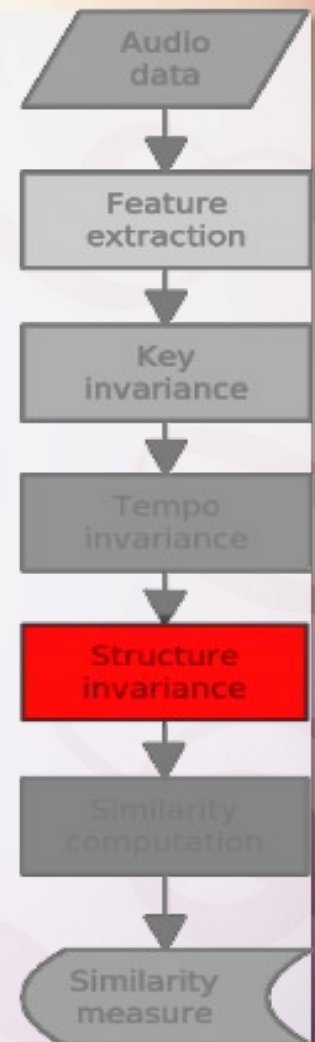
- Algorithmus zum Vergleich von zwei unterschiedlich schnellen Sequenzen





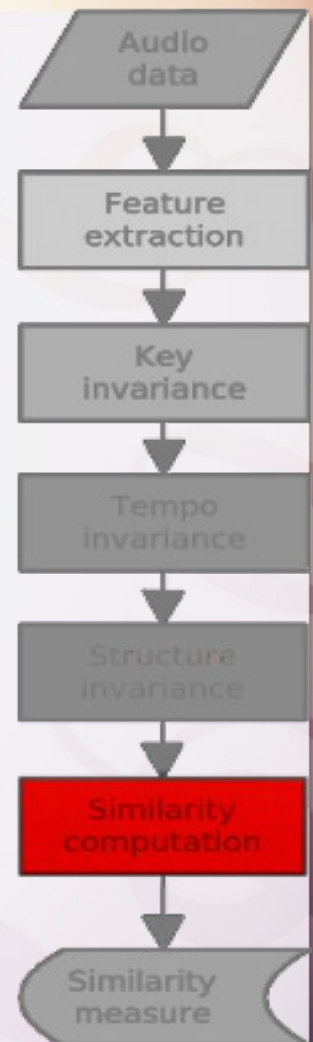
# Structure invariance

- Schlüsselfaktor bei der Bestimmung von Coversongs
- Zusammenfassung des Songs, häufig durch die meistgespieltesten Teile
- Local alignment algorithms
- Sequence Windowing



# Similarity computation

- Aufstellen eines Rankings der ähnlichsten Stücke
- Einige Aspekte wie time-warping besitzen schon eine Ähnlichkeitsberechnung
- Eher selten: Hidden-Markov-Models



**Table 2** Cover song identification methods and their ways to overcome departures from the "canonical" song. A blank space denotes no specific treatment for them. Abbreviations for extracted features are PBFV for polyphonic binary feature vector, and PCP for pitch class profile. Abbreviation for key invariance is OTI for optimal transposition index. Abbreviations for tempo invariance are DP for dynamic programming, and HMM for Hidden Markov Models. Abbreviations for similarity computation are DTW for dynamic time warping, MLSS for most likely sequence of states, and NCD for normalized compression distance.

Reference(s)	Extracted feature	Key invariance	Tempo invariance	Structure invariance	Similarity computation
Ahonen & Lemstrom [2]	Chords	Relative changes			NCD
Bello [4]	Chords	$K$ transpositions	DP		Edit distance
Egorov & Linetsky [20]	PCP	OTI	DP	DP	Match length
Ellis et al. [21, 23]	PCP	All transpositions	Beat		Cross-correlation
Foote [25]	Energy + Spectral		DP		DTW
Gómez & Herrera [28]	PCP	Key estimation	DP		DTW
Gómez et al. [29]	PCP	Key estimation	DP	Repeated patterns	DTW
Izmirli [35]	Key templates		DP		DTW
Jensen et al. [36]	PCP	All transpositions	Fourier transform		Frobenius norm
Jensen et al. [37]	PCP	2D autocorrelation	2D autocorrelation		Euclidean distance
Kim et al. [38, 39]	PCP + Delta PCP	All transpositions			Dot product
Kim & Perelstein [40]	PCP	Relative changes	HMM		MLSS
Kurth & Muller [41]	PCP	All transpositions	Temporal comp./exp.	Sequence windowing	Dot product
Lee [43]	Chords	Key estimation	DP		DTW
Marolt [49]	Melodic	Key estimation	DP	Repeated patterns	Cross-correlation
Marolt [50]	Melodic	2D spectrum	Beat + 2D spectrum	Sequence windowing	Euclidean distance
Müller et al. [53]	PCP		Temporal comp./exp.	Sequence windowing	Dot product
Nagano et al. [55]	PBFV	All transpositions	Beat + DP	Seq. windowing + DP	Match length
Sailer & Dressler [68]	Melodic		Relative changes		Edit distance
Serrà et al. [74, 76]	PCP	OTIs	DP	DP	Match length
Tsai et al. [78, 79]	Melodic	$K$ transpositions	DP		DTW
Yang [89]	Spectral		DP	Linearity filtering	Match length

# Was nun?

- i. Einleitung
- ii. Lösungsansätze
- iii. LSH**
- iv. Fazit



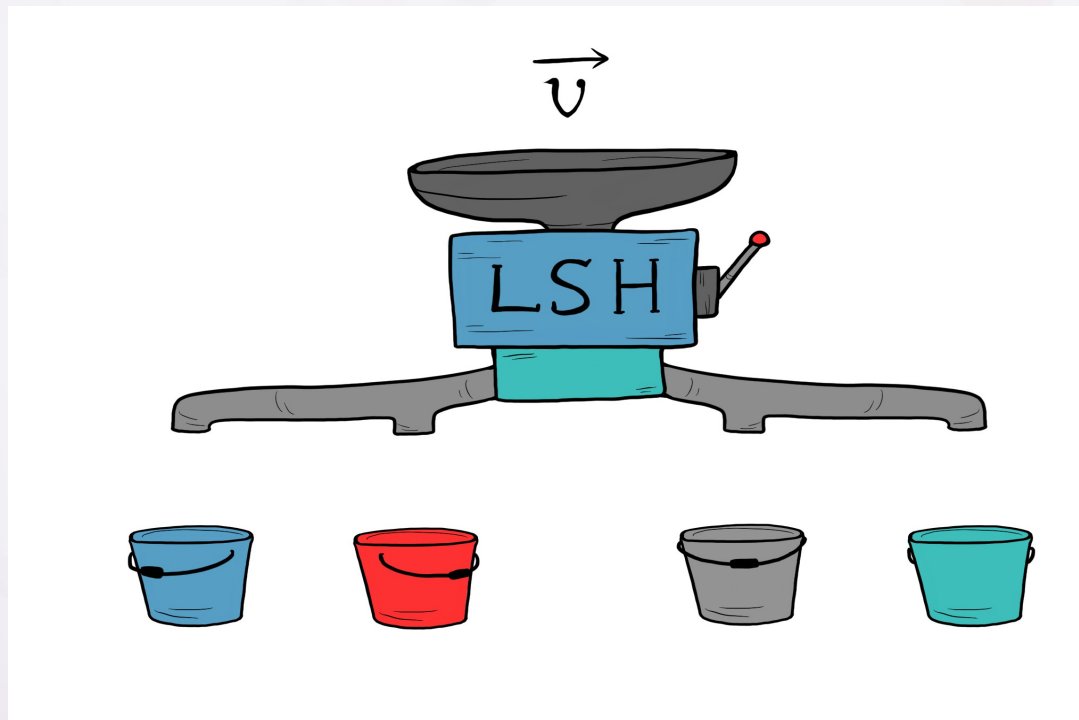
# LSH

- Locality Sensitive Hashing
- Kann ähnliche Einträge in großen Datenbanken finden
- Randomized Algorithm



# LSH

- **Idee:** Punkte, die nah beieinander sind, sind auch nach einer Projektion nah beieinander



# 1. Schritt: Skalarprodukt

Durch Gaußverteilung zufällig  
gewählter Vektor

Zufällige Variable  
zwischen 0 und w

$$h^{x,b}(\vec{v}) = \left\lfloor \frac{\vec{x} \cdot \vec{v} + b}{w} \right\rfloor$$

Breite der Eimer, in die sortiert wird

- K-fache Anwendung ergibt k-dimensionale Repräsentation
- Ergebnis: Datenpunkte in "Eimern", beschrieben durch k Integer Indizes

## 2. Schritt: Hash-Implementierung

- Konventionelles Hashing zur schnelleren Suche

$$T_1 = \left( \sum_i H_i k_i \right) \bmod P_1$$

Integer weights

Größe der Hash-Table

- Zweifaches Hashen um zufällige Übereinstimmungen zu vermeiden

# Was haben wir gelernt?

- Cover Song Recognition besteht aus mehreren Schritten
- Jeder Schritt besteht aus komplexen Berechnungen

# Was nun?

- i. Einleitung
- ii. Lösungsansätze
- iii. LSH
- iv. Fazit**



# Fazit

- Evaluationen meist nur in zusammengefasster Form
- **Problem:** einzelne Teilschritte besitzen keine eigene Analyse

# MIREX: Cover Song Identification

- Sammlung von 1000 Musikstücken
- 30 Coversets mit jeweils 11 verschiedenen Covern eines Songs
- **Aufgabe:** Nach Übergabe eines der Coversongs Finden der 10 anderen

# MIREX Evaluation

Jahr	Teiln.	Durchschnitt	Bester	Schlechtester
2007	8	21%	52%	2%
2008	8	50%	75%	23%
2009	3	53%	75%	20%
2010	3	23%	29%	15%
2011	2	17%	21%	13%
2013	2	40%	52%	28%
2014	2	2%	2%	2%
2015	2	30%	41%	19%
2016	5	21%	58%	0%

# Danke für eure Aufmerksamkeit!



## Gibt es noch Fragen?

# Quellen

- Fujishima, Takuya. "Realtime chord recognition of musical sound: A system using common lisp music." Proc. ICMC. Vol. 1999. 1999.
- Slaney, Malcolm, and Michael Casey. "Locality-sensitive hashing for finding nearest neighbors [lecture notes]." IEEE Signal Processing Magazine 25.2 (2008): 128-131.
- Marolt, Matija. "A mid-level representation for melody-based retrieval in audio collections." IEEE Transactions on Multimedia 10.8 (2008): 1617-1625.
- Serra, Joan, Emilia Gómez, and Perfecto Herrera. "Audio cover song identification and similarity: background, approaches, evaluation, and beyond." Advances in Music Information Retrieval. Springer Berlin Heidelberg, 2010. 307-332.
- Jensen, Jesper Højvang, et al. "A chroma-based tempo-insensitive distance measure for cover song identification." Third Music Information Retrieval Evaluation eXchange (2007).



# Quellen

- Casey, Michael, and Malcolm Slaney. "Fast recognition of remixed music audio." 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07. Vol. 4. IEEE, 2007.
- Poliner, Graham E., et al. "Melody transcription from music audio: Approaches and evaluation." IEEE Transactions on Audio, Speech, and Language Processing 15.4 (2007): 1247-1256.
- [http://www.music-ir.org/mirex/wiki/MIREX\\_HOME](http://www.music-ir.org/mirex/wiki/MIREX_HOME)
- <https://xkcd.com/359/>
- <https://pixabay.com/de/play-abspielen-musik-filme-tragbar-468292/>