



Hochschule **RheinMain**
University of Applied Sciences
Wiesbaden Rüsselsheim

MUSIC IDENTIFICATION USING INVARIANT SPECTROGRAM FEATURES

Fachseminar “Machine Learning”

Letztes Update: 13. Februar 2016

Marcus Michaely

Studienbereich Informatik
Hochschule RheinMain



GLIEDERUNG

1. Einleitung
2. Grundlagen
3. Algorithmus
4. Diskussion
5. Fazit
6. Quellen
7. Fragen

EINLEITUNG

WOMIT HABE ICH MICH BESCHÄFTIGT

- ▶ **Mit welchem Paper habe ich mich beschäftigt?**

WOMIT HABE ICH MICH BESCHÄFTIGT

- ▶ **Mit welchem Paper habe ich mich beschäftigt?**
An Industrial Strength Audio Search Algorithm von Avery Wang im Jahre 2003 [1]

WOMIT HABE ICH MICH BESCHÄFTIGT

- ▶ **Mit welchem Paper habe ich mich beschäftigt?**
An Industrial Strength Audio Search Algorithm von Avery Wang im Jahre 2003 [1]
- ▶ **Worum geht es in dem Paper?**

WOMIT HABE ICH MICH BESCHÄFTIGT

- ▶ **Mit welchem Paper habe ich mich beschäftigt?**
An Industrial Strength Audio Search Algorithm von Avery Wang im Jahre 2003 [1]
- ▶ **Worum geht es in dem Paper?**
Funktionsweise des Musikerkennungsdienstes Shazam

WOMIT HABE ICH MICH BESCHÄFTIGT

- ▶ **Mit welchem Paper habe ich mich beschäftigt?**
An Industrial Strength Audio Search Algorithm von Avery Wang im Jahre 2003 [1]
- ▶ **Worum geht es in dem Paper?**
Funktionsweise des Musikerkennungsdienstes Shazam
- ▶ **Wie funktioniert das im Groben?**

WOMIT HABE ICH MICH BESCHÄFTIGT

- ▶ **Mit welchem Paper habe ich mich beschäftigt?**
An Industrial Strength Audio Search Algorithm von Avery Wang im Jahre 2003 [1]
- ▶ **Worum geht es in dem Paper?**
Funktionsweise des Musikerkennungsdienstes Shazam
- ▶ **Wie funktioniert das im Groben?**
Zu erkennendes Musikstück wird mit Mobiltelefon aufgenommen und der Musikerkennungsdienst liefert den Namen des aufgenommenen Songs.

WOMIT HABE ICH MICH BESCHÄFTIGT

- ▶ **Mit welchem Paper habe ich mich beschäftigt?**
An Industrial Strength Audio Search Algorithm von Avery Wang im Jahre 2003 [1]
- ▶ **Worum geht es in dem Paper?**
Funktionsweise des Musikerkennungsdienstes Shazam
- ▶ **Wie funktioniert das im Groben?**
Zu erkennendes Musikstück wird mit Mobiltelefon aufgenommen und der Musikerkennungsdienst liefert den Namen des aufgenommenen Songs.
- ▶ **Im Anschluss des Vortrags sollte klar sein?**

WOMIT HABE ICH MICH BESCHÄFTIGT

- ▶ **Mit welchem Paper habe ich mich beschäftigt?**
An Industrial Strength Audio Search Algorithm von Avery Wang im Jahre 2003 [1]
- ▶ **Worum geht es in dem Paper?**
Funktionsweise des Musikerkennungsdienstes Shazam
- ▶ **Wie funktioniert das im Groben?**
Zu erkennendes Musikstück wird mit Mobiltelefon aufgenommen und der Musikerkennungsdienst liefert den Namen des aufgenommenen Songs.
- ▶ **Im Anschluss des Vortrags sollte klar sein?**
 - ▶ Wie man ein Spektrogramm einer Audioaufnahme interpretieren kann.

WOMIT HABE ICH MICH BESCHÄFTIGT

- ▶ **Mit welchem Paper habe ich mich beschäftigt?**
An Industrial Strength Audio Search Algorithm von Avery Wang im Jahre 2003 [1]
- ▶ **Worum geht es in dem Paper?**
Funktionsweise des Musikerkennungsdienstes Shazam
- ▶ **Wie funktioniert das im Groben?**
Zu erkennendes Musikstück wird mit Mobiltelefon aufgenommen und der Musikerkennungsdienst liefert den Namen des aufgenommenen Songs.
- ▶ **Im Anschluss des Vortrags sollte klar sein?**
 - ▶ Wie man ein Spektrogramm einer Audioaufnahme interpretieren kann.
 - ▶ Was invariante Spektrogrammmerkmale sind.

WOMIT HABE ICH MICH BESCHÄFTIGT

- ▶ **Mit welchem Paper habe ich mich beschäftigt?**

An Industrial Strength Audio Search Algorithm von Avery Wang im Jahre 2003 [1]

- ▶ **Worum geht es in dem Paper?**

Funktionsweise des Musikerkennungsdienstes Shazam

- ▶ **Wie funktioniert das im Groben?**

Zu erkennendes Musikstück wird mit Mobiltelefon aufgenommen und der Musikerkennungsdienst liefert den Namen des aufgenommenen Songs.

- ▶ **Im Anschluss des Vortrags sollte klar sein?**

- ▶ Wie man ein Spektrogramm einer Audioaufnahme interpretieren kann.
- ▶ Was invariante Spektrogrammmerkmale sind.
- ▶ Wie der Algorithmus sich diese beiden Punkte zunutze macht.

GRUNDLAGEN

INVARIANT

Definition von invariant laut Duden?

INVARIANT

Definition von invariant laut Duden?

Invariant Adjektiv = bei veränderten Bedingungen unverändert bleiben.[2]

SPEKTROGRAMM

Was ist ein Spektrogramm?

SPEKTROGRAMM

Was ist ein Spektrogramm?

so sieht ein Spektrogramm aus

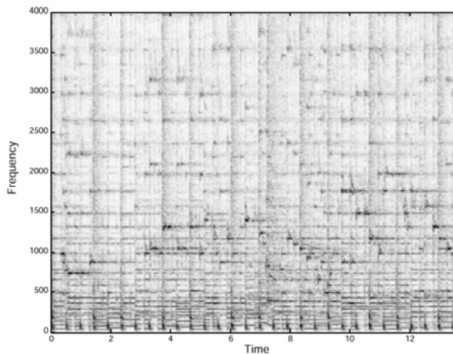


Abbildung: Spektrogramm [1]

SPEKTROGRAMM

Was ist ein Spektrogramm?

so sieht ein Spektrogramm aus

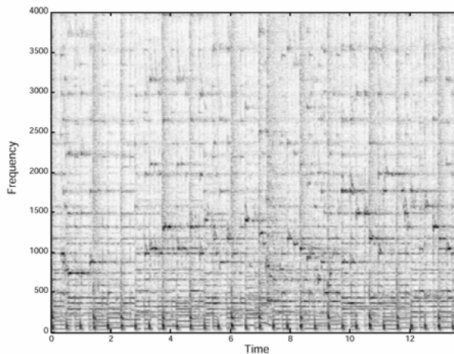


Abbildung: Spektrogramm [1]

SINUS

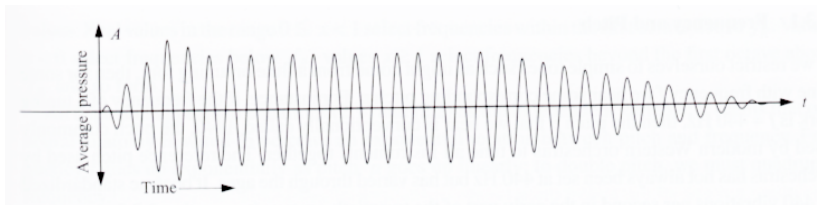


Abbildung: Sinus[3]

- Amplitude

SINUS

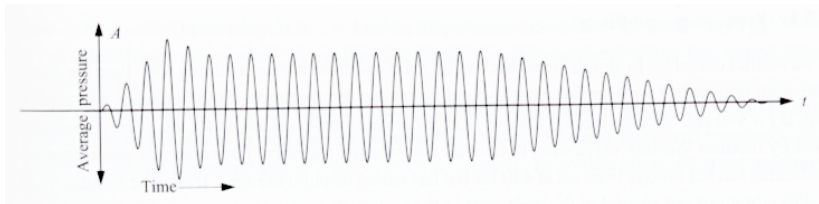


Abbildung: Sinus[3]

- ▶ Amplitude = Lautstärke
- ▶ Periodendauer

SINUS

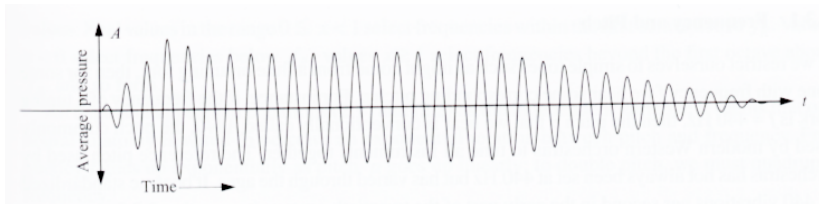


Abbildung: Sinus[3]

- ▶ Amplitude = Lautstärke
- ▶ Periodendauer = Tonhöhe

ÜBERLAGERUNG VON SINUSSCHWINGUNGEN

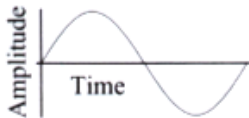


Abbildung: Sinus [3]

ÜBERLAGERUNG VON SINUSSCHWINGUNGEN

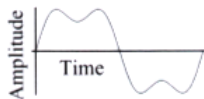
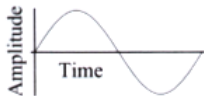


Abbildung: Sinus [3]

ÜBERLAGERUNG VON SINUSSCHWINGUNGEN

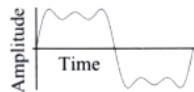
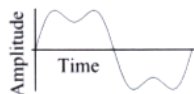
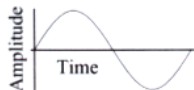


Abbildung: Sinus [3]

ÜBERLAGERUNG VON SINUSSCHWINGUNGEN

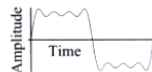
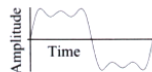
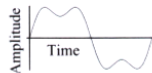
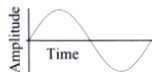


Abbildung: Sinus [3]

FOURIER GRAFISCH

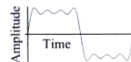
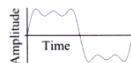
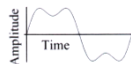
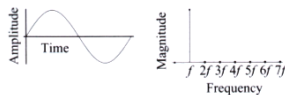


Abbildung: Sinus [3]

FOURIER GRAFISCH

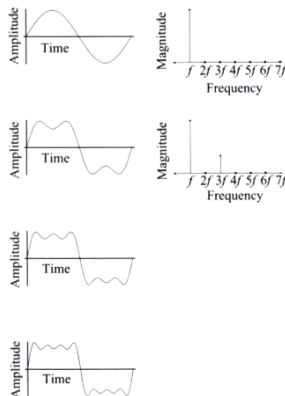


Abbildung: Bildmontage Sinus [3]

FOURIER GRAFISCH

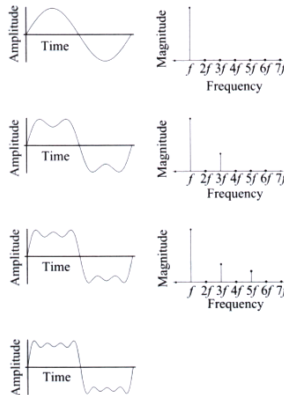


Abbildung: Sinus [3]

FOURIER GRAFISCH

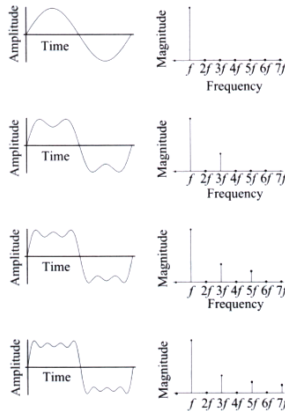


Abbildung: Sinus [3]

FOURIER FORMEL

Fourier integral

$$F(j\omega) = \int_{-\infty}^{\infty} f(t) \cdot e^{-j\omega t} dt$$

[4]

VON FOURIER ZUM SPEKTROGRAMM

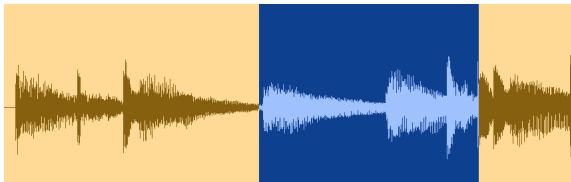


Abbildung: Screenshot Transcribe [5]

VON FOURIER ZUM SPEKTROGRAMM

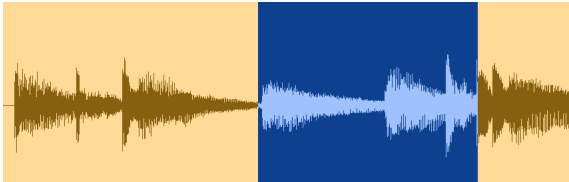


Abbildung: Screenshot Transcribe [5]

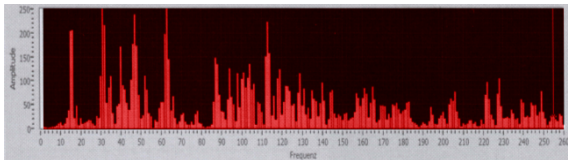


Abbildung: Audioprogramm Fourier [6]

VON FOURIER ZUM SPEKTROGRAMM

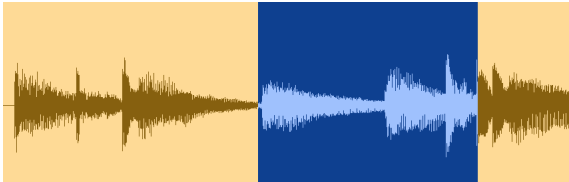


Abbildung: Screenshot Transcribe [5]

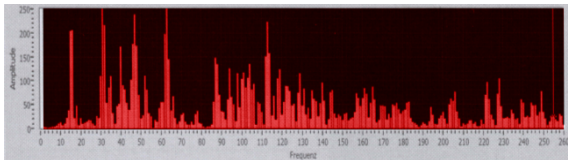


Abbildung: Audioprogramm Fourier [6]

VON FOURIER ZUM SPEKTROGRAMM

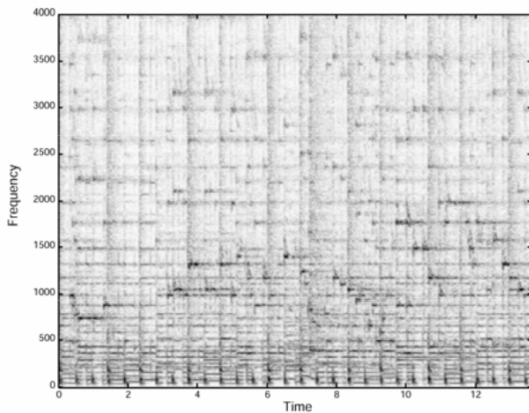


Abbildung: Spektrogramm [1]

VON FOURIER ZUM SPEKTROGRAMM

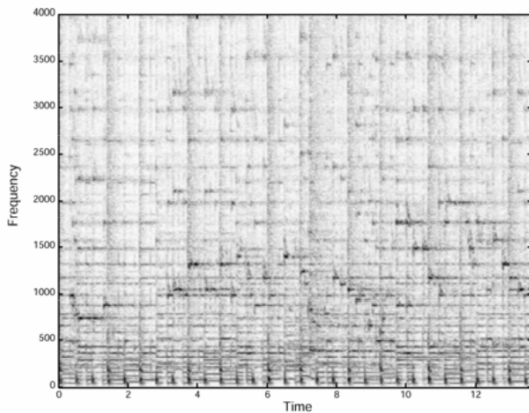


Abbildung: Spektrogramm [1]

DARSTELLUNG MONOPHONER TON IM SPEKTROGRAMM

DARSTELLUNG MONOPHONER TON IM SPEKTROGRAMM

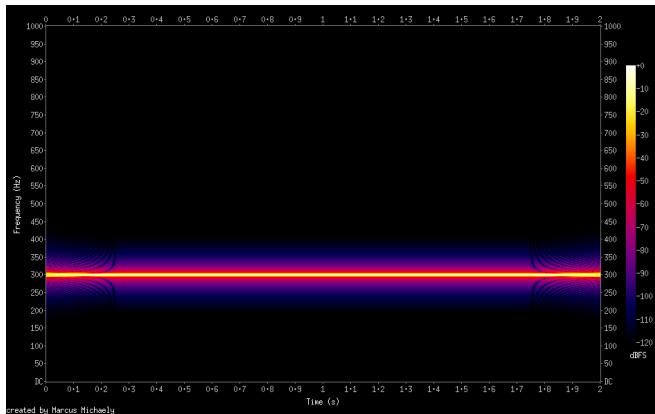


Abbildung: monophoner Ton

play sound

DARSTELLUNG RHYTHMUS IM SPEKTROGRAMM

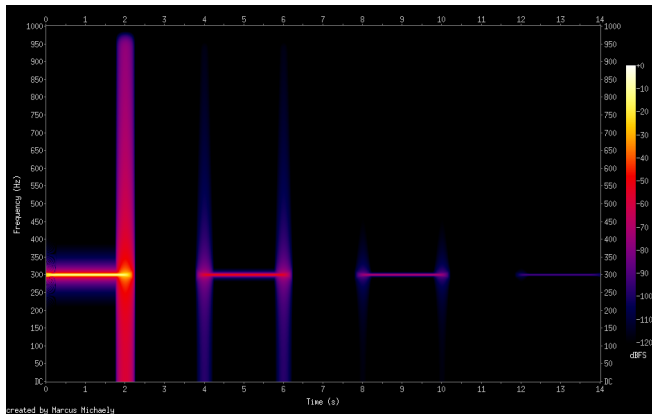


Abbildung: Rhythmus Lautstärke nimmt ab

play sound

POLYPHONER KLANG IM SPEKTROGRAMM

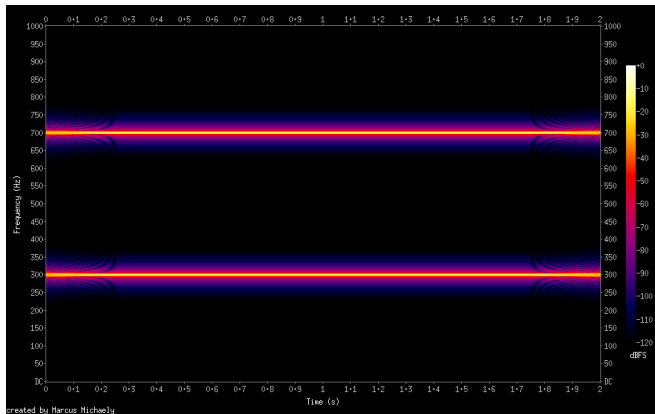


Abbildung: polyphoner Klang

play sound

TONLEITER IM SPEKTROGRAMM

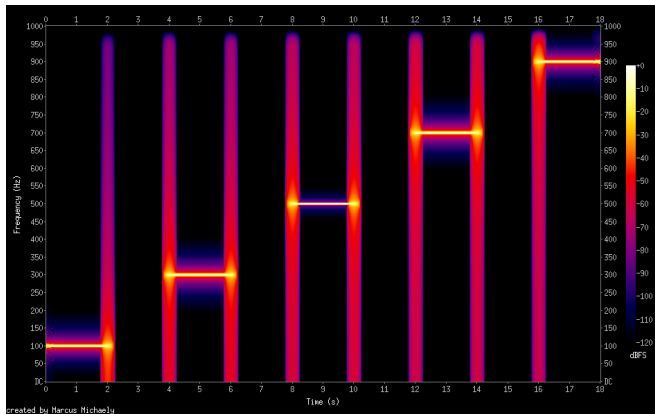


Abbildung: Tonleiter

play sound

WIR VERLASSEN DIE OPTIMALE AUDIOWELT

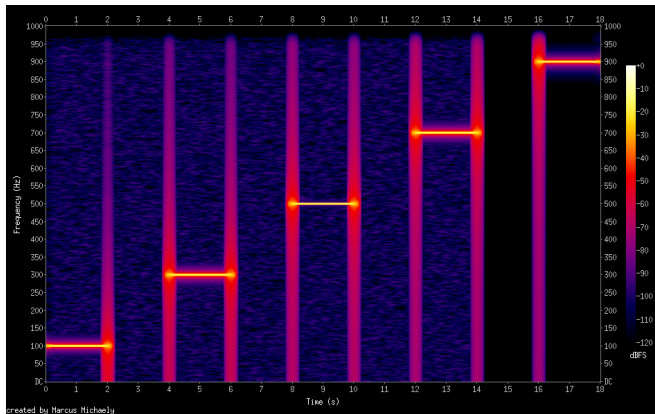


Abbildung: Tonleiter mit Hintergrundrauschen

play sound

WIR VERLASSEN DIE OPTIMALE AUDIOWELT

- Verzerrung

WIR VERLASSEN DIE OPTIMALE AUDIOWELT

- Verzerrung
- Clipping (Abschneiden des Signals)

WIR VERLASSEN DIE OPTIMALE AUDIOWELT

- Verzerrung
- Clipping (Abschneiden des Signals)
- Audiokomprimierung (GSM, MP3 ..)

WIR VERLASSEN DIE OPTIMALE AUDIOWELT

- Verzerrung
- Clipping (Abschneiden des Signals)
- Audiokomprimierung (GSM, MP3 ..)

ALGORITHMUS

ALGORITHMUS STEP BY STEP

ALGORITHMUS STEP BY STEP

I) Spektrogramm erstellen

SPEKTROGRAMM EINER REALEN TESTAUFNAHME

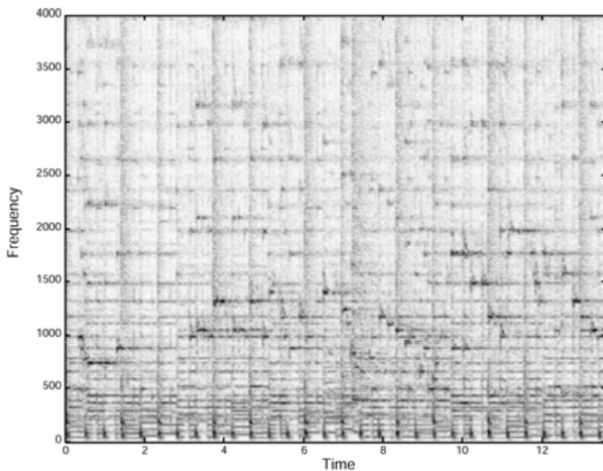


Abbildung: Spektrogramm [1]

VON SPEKTROGRAMM ZUR CONSTELLATION MAP

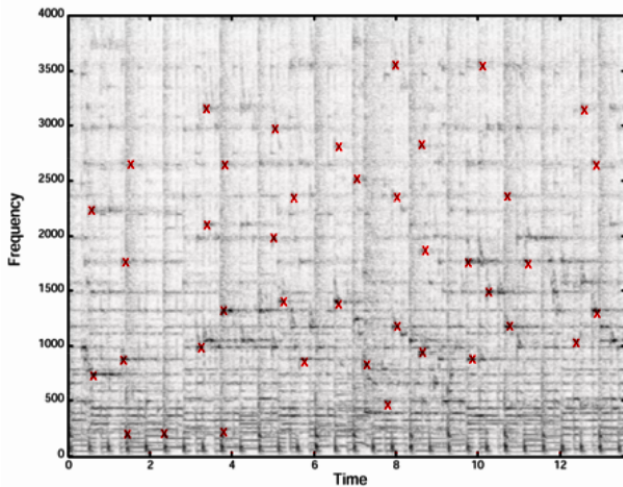


Abbildung: Bildmontage aus Spektrogramm [1]

ALGORITHMUS STEP BY STEP

I) Spektrogramm erstellen

ALGORITHMUS STEP BY STEP

I) Spektrogramm erstellen

II) **Constellation Map erstellen (Invariant: Lautstärke)**

VON SPEKTROGRAMM ZUR CONSTELLATION MAP

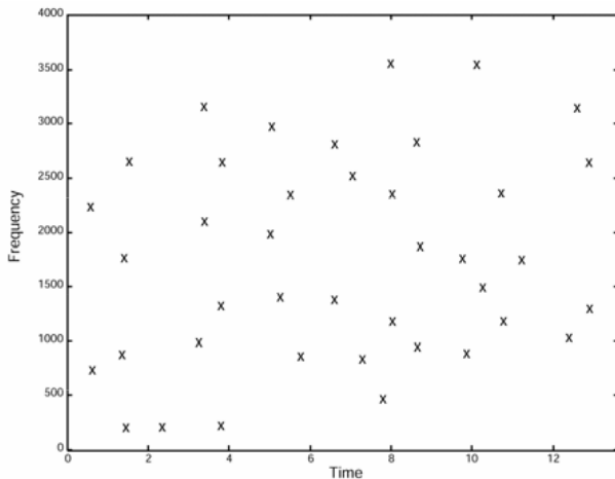


Abbildung: Constellation Map [1]

ALGORITHMUS STEP BY STEP

- I) Spektrogramm erstellen
- II) Constellation Map erstellen (Invariant: Lautstärke)

ALGORITHMUS STEP BY STEP

- I) Spektrogramm erstellen
- II) Constellation Map erstellen (Invariant: Lautstärke)
- III) **Fast Combinatorial Hashes (Translation bzw. Time-Invariant)**

FAST COMBINATIONAL HASHING

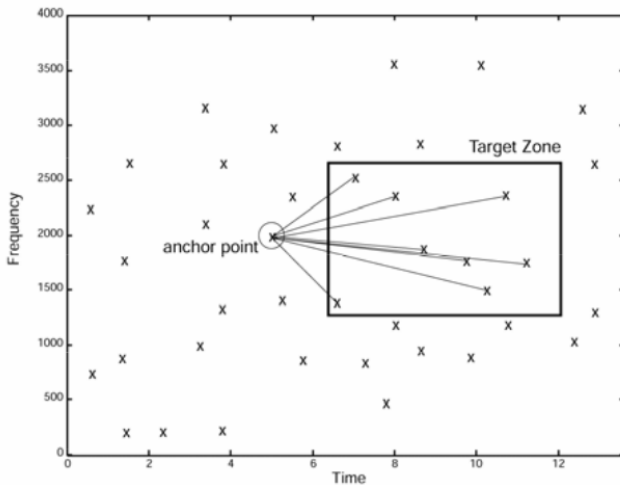
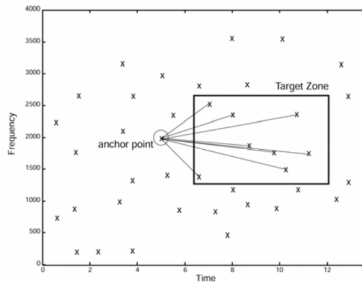


Abbildung: Constellation Map [1]

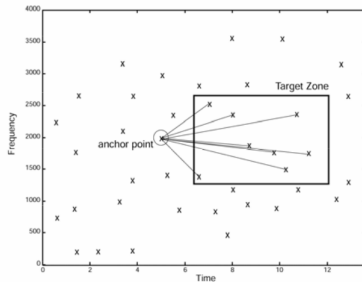
FAST COMBINATIONAL HASHING



► Anchor Point

Abbildung: Constellation Map [1]

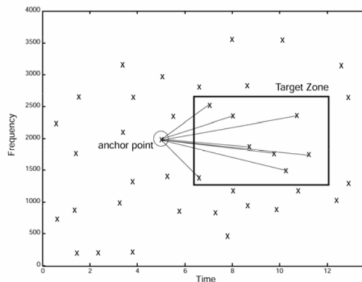
FAST COMBINATIONAL HASHING



- Anchor Point
- Target Zone

Abbildung: Constellation Map [1]

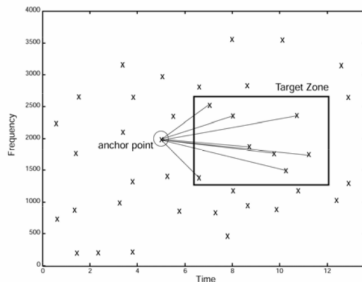
FAST COMBINATIONAL HASHING



- Anchor Point
- Target Zone
- Fan Out

Abbildung: Constellation Map [1]

FAST COMBINATIONAL HASHING



- Anchor Point
- Target Zone
- Fan Out

Abbildung: Constellation Map [1]

HASHING DETAILS

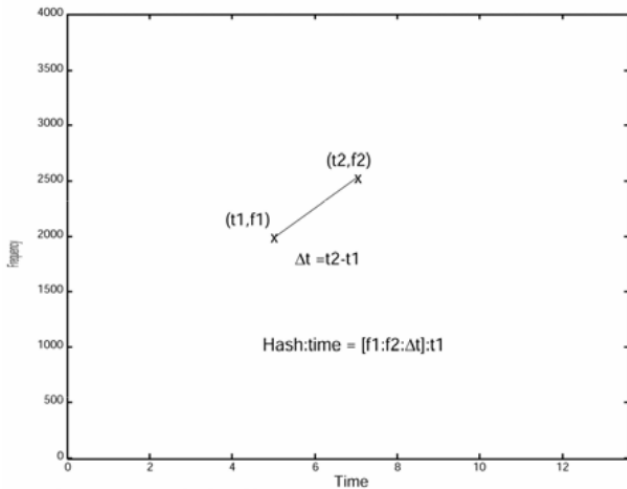
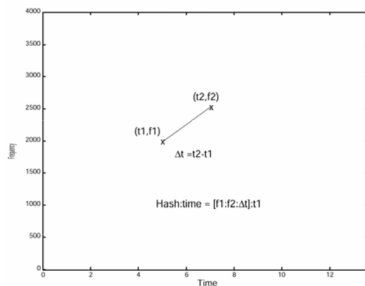


Abbildung: Hash details [1]

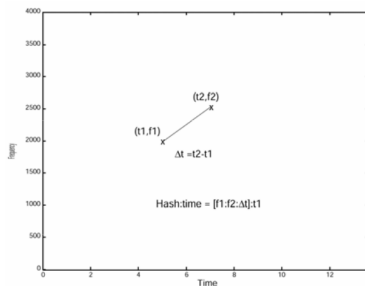
FAST COMBINATIONAL HASHING



- translation invariant also unabhängig von der Zeit

Abbildung: Hash details [1]

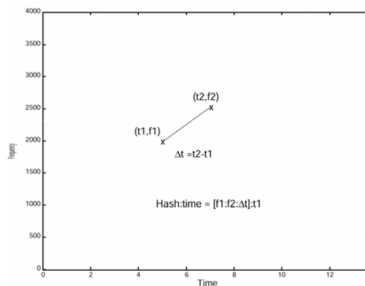
FAST COMBINATIONAL HASHING



- translation invariant also unabhängig von der Zeit
- $t1$

Abbildung: Hash details [1]

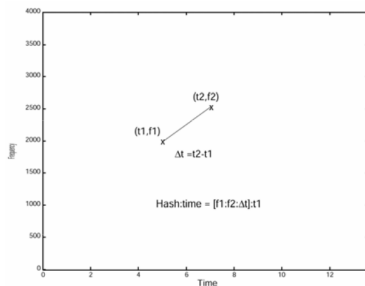
FAST COMBINATIONAL HASHING



- translation invariant also unabhängig von der Zeit
- $t1$
- $t2$

Abbildung: Hash details [1]

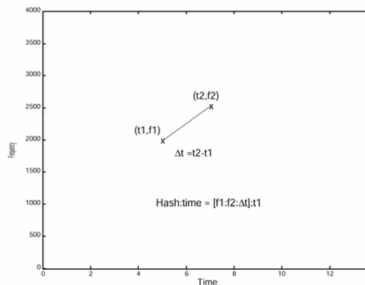
FAST COMBINATIONAL HASHING



- ▶ translation invariant also unabhängig von der Zeit
- ▶ t_1
- ▶ t_2
- ▶ Δt

Abbildung: Hash details [1]

FAST COMBINATIONAL HASHING



- ▶ translation invariant also unabhängig von der Zeit
- ▶ $t1$
- ▶ $t2$
- ▶ Δt
- ▶ $f1:f2: \Delta t$

Abbildung: Hash details [1]

ALGORITHMUS STEP BY STEP

I) Spektrogramm erstellen

II) Constellation Map erstellen (Invariant: Lautstärke)

III) Fast Combinatorial Hashes (Translation bzw. Time-Invariant)

IV) **Suche**

SUCHALGORITHMUS STEP BY STEP

- ▶ 1) Datenbank von Orginalsongs erstellen.

SUCHALGORITHMUS STEP BY STEP

- ▶ 1) Datenbank von Orginalsongs erstellen.
 - ▶ jeder Hash speichern.

SUCHALGORITHMUS STEP BY STEP

- ▶ 1) Datenbank von Originalsongs erstellen.
 - ▶ jeder Hash speichern.
 - ▶ Position des Hashes im Originalsongs

SUCHALGORITHMUS STEP BY STEP

- ▶ 1) Datenbank von Originalsongs erstellen.
 - ▶ jeder Hash speichern.
 - ▶ Position des Hashes im Originalsongs
 - ▶ Name des Originalsongs

SUCHALGORITHMUS STEP BY STEP

- ▶ I) Datenbank von Originalsongs erstellen.
 - ▶ jeder Hash speichern.
 - ▶ Position des Hashes im Originalsongs
 - ▶ Name des Originalsongs
- ▶ II) Aufnahme per Telefon machen und Hashes erstellen.

SUCHALGORITHMUS STEP BY STEP

- ▶ I) Datenbank von Originalsongs erstellen.
 - ▶ jeder Hash speichern.
 - ▶ Position des Hashes im Originalsongs
 - ▶ Name des Originalsongs
- ▶ II) Aufnahme per Telefon machen und Hashes erstellen.
- ▶ III) Datenbankabfrage welche Songs die Menge der Telefon-Hashes enthalten (Order by Value)

SUCHALGORITHMUS STEP BY STEP

- ▶ I) Datenbank von Originalsongs erstellen.
 - ▶ jeder Hash speichern.
 - ▶ Position des Hashes im Originalsongs
 - ▶ Name des Originalsongs
- ▶ II) Aufnahme per Telefon machen und Hashes erstellen.
- ▶ III) Datenbankabfrage welche Songs die Menge der Telefon-Hashes enthalten (Order by Value)
- ▶ IV) **Achtung es die Hashes sind noch nicht in der richtigen Reihenfolge!**

NICHT ÜBEREINSTIMMUNG

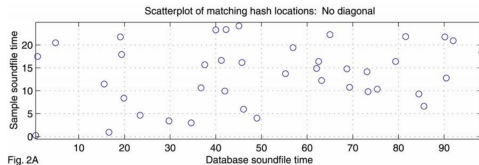


Abbildung: Punktwolke [1]

NICHT ÜBEREINSTIMMUNG

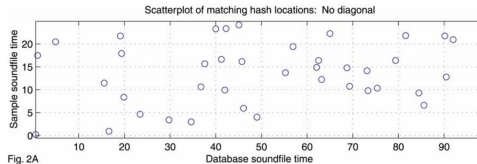


Abbildung: Punktwolke [1]

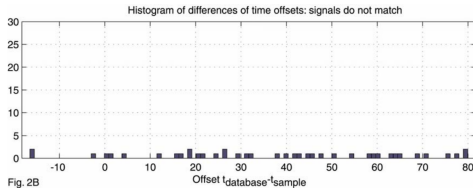


Abbildung: Histogram [1]

ÜBEREINSTIMMUNG

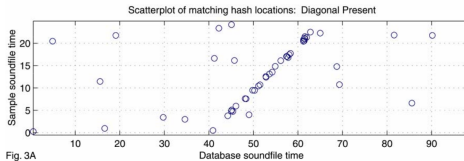


Abbildung: Punktwolke [1]

ÜBEREINSTIMMUNG

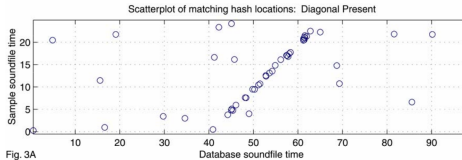


Abbildung: Punktwolke [1]

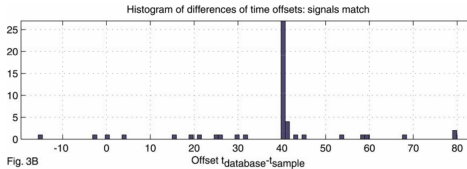


Abbildung: Histogram [1]

DISKUSSION

WAS WAR INVARIANT

- Amplitudeninvariant Amplitude fällt bei Constellation Map weg

WAS WAR INVARIANT

- Amplitudeninvariant Amplitude fällt bei Constellation Map weg
- Translationsinvariant zeitliche Komponente fällt durch combinatorial hashing weg.

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen:

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm:

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja
- Ausfälle:

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja
- Ausfälle: ja (in geringem Ausmaß)

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja
- Ausfälle: ja (in geringem Ausmaß)
- Andere Musik:

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja
- Ausfälle: ja (in geringem Ausmaß)
- Andere Musik: ja

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja
- Ausfälle: ja (in geringem Ausmaß)
- Andere Musik: ja
- variabler Aufnahmezeitpunkt der Testaufnahme:

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja
- Ausfälle: ja (in geringem Ausmaß)
- Andere Musik: ja
- variabler Aufnahmezeitpunkt der Testaufnahme: egal weil translationsinvariant

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja
- Ausfälle: ja (in geringem Ausmaß)
- Andere Musik: ja
- variabler Aufnahmezeitpunkt der Testaufnahme: egal weil translationsinvariant
- Schneller bzw. langsamer Abspielen:

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja
- Ausfälle: ja (in geringem Ausmaß)
- Andere Musik: ja
- variabler Aufnahmezeitpunkt der Testaufnahme: egal weil translationsinvariant
- Schneller bzw. langsamer Abspielen: nein

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja
- Ausfälle: ja (in geringem Ausmaß)
- Andere Musik: ja
- variabler Aufnahmezeitpunkt der Testaufnahme: egal weil translationsinvariant
- Schneller bzw. langsamer Abspielen: nein
- Tonhöhenänderung:

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja
- Ausfälle: ja (in geringem Ausmaß)
- Andere Musik: ja
- variabler Aufnahmezeitpunkt der Testaufnahme: egal weil translationsinvariant
- Schneller bzw. langsamer Abspielen: nein
- Tonhöhenänderung: nein

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja
- Ausfälle: ja (in geringem Ausmaß)
- Andere Musik: ja
- variabler Aufnahmezeitpunkt der Testaufnahme: egal weil translationsinvariant
- Schneller bzw. langsamer Abspielen: nein
- Tonhöhenänderung: nein
- Coversongs/Livemusik:

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja
- Ausfälle: ja (in geringem Ausmaß)
- Andere Musik: ja
- variabler Aufnahmezeitpunkt der Testaufnahme: egal weil translationsinvariant
- Schneller bzw. langsamer Abspielen: nein
- Tonhöhenänderung: nein
- Coversongs/Livemusik: nein

STÖRUNGEN DIE DER ALGORITHMUS VERKRAFTEN KANN

- Stimmen: ja
- Verkehrslärm: ja
- Ausfälle: ja (in geringem Ausmaß)
- Andere Musik: ja
- variabler Aufnahmezeitpunkt der Testaufnahme: egal weil translationsinvariant
- Schneller bzw. langsamer Abspielen: nein
- Tonhöhenänderung: nein
- Coversongs/Livemusik: nein
- False Positives weil Song im Vordergrund eine niedrigere Merkmaldichte hat als der Song im Hintergrund (Vordergrund Klassik(Streicher) Hintergrund HipHop(Beats))

FAZIT

Algorithmus ist gut um mit Hilfe von kurzen Testaufnahmen welche mit Störsignalen verunreinigt wurden die entsprechenden Original-Musikstücke in einer großen Datenbank zu finden. Laut Wang 20000 Lieder mit einer Suchzeit von 5-500 Millisekunden.

QUELLEN



Avery Li-Chun Wang.

An industrial-strength audio search algorithm.

October 2003.



Duden eintrag: Invariant.

<http://www.duden.de/rechtschreibung/invariant>.

Accessed: 2015-11-30.



Musimathics musimathics vol.1.

www.mitpress.mit.edu.

Accessed: 2016-01-16.



Wikipedia eintrag: Joseph fourier.

https://de.wikipedia.org/wiki/Joseph_Fourier.

Accessed: 2015-11-30.



Transcribe musiksoftware.

www.sevenstring.com.

Accessed: 2016-01-16.



Audiospektrum-Analyse mit Mikrokontrollern felix irmscher.

www.franzis.de.

FRAGEN