

## README

### Aufbau der Ablage

Die Skripte in der Ablage sind nach den verschiedenen Analysearten sortiert. Es sind nur die im Rahmen dieser Arbeit entstandenen Programme in der Sprache R hochgeladen, die genutzten Datensätze sind verlinkt. Eine Anleitung, um die Programme selbst lokal laufen zu lassen findet sich in diesem Dokument im Abschnitt 'Programme selbst laufen lassen'. Die Ablage ist wie folgt aufgebaut:

Name	Analyseart	Beschreibung
3-Benennung	-	Umbenennung der Dateien aus den Datensätzen von Neuwirth et al. und Hentschel et al. damit diese von den Programmen dieser Arbeit eingelesen werden können. Diese Programme müssen zu Anfang einmal ausgeführt werden.
4-Datenbereinigung	-	Bereinigung der Tabellen aus den Datensätzen von Neuwirth et al. und Hentschel et al. bzgl. Format und Inhalt damit diese von den Programmen dieser Arbeit eingelesen werden können. Diese Programme sind jeweils – sofern nötig – in die anderen integriert.
ana1-UG-Rang-Häufigkeit	Unigramm	a. Untersuchung der Rang-Häufigkeit b. Vergleich der R-H mit Mandelbrot-Zipf-Kurve c. Vergleich der R-H verschiedener Datensätze
ana2-BG-Übergänge	Bigramm	a. Häufigkeit der Übergänge zwischen den 25 häufigsten Akkorden in einem Datensatz b. Vergleich zweier Datensätze
ana3-BG-Entropie	Bigramm	a. Vgl. der Entropie einer Akkordstichprobe aus allen Akkorden mit Akkorden mit bestimmter Eigenschaft (aus demselben Datensatz) b. Vgl. der Entropie von Akkordstichproben eines Datensatzes mit Akkorden eines anderen 1. Normal, Differenzierung aller Akkordtoken (wie in ana3a) 2. Reduktion auf Stufen (I, II, ..)
ana4-BG-Symmetrie	Bigramm	a. Untersuchung der Symmetrie von Akkordübergängen (Bigrammen)

### Programme selbst laufen lassen

Um die Programme selbst auf dem lokalen Computer laufen zu lassen, muss wie folgt vorgegangen werden:

1. Herunterladen der Datensätze von Neuwirth et al. und Hentschel et al. über Github
  - a. Der Datensatz der annotierten Beethoven Streicherquartette von Neuwirth et al. ist unter folgendem Link verfügbar: <https://github.com/DCMLab/ABC.git>
  - b. Der Datensatz der annotierten Beethoven Klaviersonaten von Hentschel et al. ist unter folgendem Link verfügbar: [https://github.com/DCMLab/beethoven\\_piano\\_sonatas.git](https://github.com/DCMLab/beethoven_piano_sonatas.git)

Folgende Dateien aus den Datensätzen werden benötigt:

- Alle .tsv Tabellen aus dem Datensatz (a.) unter ABC/harmonies/ bzw. beethoven\_piano\_sonatas/harmonies/

Es können beliebige Daten verwendet werden, sofern die Annotationsform und Name passt.

Die Tabellen müssen, damit sie von den Programmen erkannt und eingelesen werden können, wie folgt benannt werden:

[KOM]-[A][xx]-M[x].tsv

wobei KOM für das Komponistenkürzel steht (LVB für L. v. Beethoven, WAM für W. A. Mozart), A für die Stückart (S für Sonate, Q für Streicherquartett) mit der üblichen Nummerierung (xx aus 00-99) und M für den Satz mit Nummer (x aus 0-9)

Die Datei zum dritten Satz der ersten Beethoven-sonate heisst also: LVB-S01-M3.tsv

Programme zur Umbenennung finden sich unter 3-Benennung.

2. Herunterladen der Dateien von diesem Projekt, diese sind auf Github unter <https://github.com/rahulandrea/rolling-over-beethoven.git> verfügbar. Eine Übersicht aller Programme findet sich in diesem Dokument im Abschnitt 'Aufbau der Ablage'.
3. Die Programme von diesem Projekt sind in der Sprache R. Es wird die Entwicklungsumgebung RStudio empfohlen. Genauere Anwendungshinweise zu den Programmen finden sich in den jeweiligen Berichten.

Damit die Programme auf die Datensätze zugreifen können, muss das Arbeitsverzeichnis für R entsprechend festgelegt werden. Die Einsicht des aktuellen Arbeitsverzeichnisses ist über die in RStudio integrierte Shell mit folgendem Befehl möglich:

```
> getwd()
```

Das Ändern des Arbeitsverzeichnisses ist mit folgendem Befehl möglich:

```
> setwd("/DCMLab-beethoven_piano_sonatas-fl31d6d/harmonies")
```

Im Arbeitsverzeichnis sollten zwei Unterverzeichnisse angelegt werden; eines in welchem die Tabellen aus dem Datensatz liegen und eines in welchem die Outputs der Programme gespeichert werden. Die Unterverzeichnisse sollten Daten bzw. Ergebnisse genannt werden, da in den Skripten Verweise über diesen Pfad laufen.

4. Es wird weiter empfohlen dieselbe Ordnerstruktur zu behalten, da es Verweise auf andere Dateien gibt, die sonst nicht verstanden würden.

Bspw. in der Datei ana1a-UG-rank-freq.R

```
8 source("Vers-B/4-Datenbereinigung/4a-combined-data.R")
```