Audio Feature Extraction for Corpus Analysis

Anja Volk
Sound and Music Technology

5 Dec 2017
Corpus analysis

- What is corpus analysis
  - study a large corpus of music for gaining insights on general trends with computational methods

- Why?
  - MIR: provide access to large corpora of music
  - Musicology: research music from a data-rich perspective
    - Test musicological hypotheses

- Today: corpus analysis of audio features on
  - choruses
  - hooks
Recap

- Automatic Segmentation of music: Applications?
  - E.g. games, indexing for search in large collections, most salient part
- Automatic Segmentation of music: what cues do humans use?
  - Gaps/change in musical features
  - Repetition
  - Closure
  - ...
- Computational approaches to segmentation
  - Local gaps: Local boundary detection (LBDM)
  - Expectation: Information-theoretic approaches
  - Rule-based vs. data-driven models
Today: Corpus analysis

- in language studies: corpus linguistics
- in musicology:
  - statistical musicology
  - data-driven musicology
  - empirical musicology
  - ...
- examples:
  - Syncopation patterns in ragtime
    - See lecture on Rhythm and Meter in SMT course
  - Huron: the melodic arch
  - Rodriguez-Zivic: perception & musical style
Corpus analysis

- in language studies: corpus linguistics
- in musicology:
  - statistical musicology
  - data-driven musicology
  - empirical musicology
  - ...
- examples:
  - Syncopation patterns in ragtime
    - See lecture on Rhythm and Meter in SMT course
  - Huron: the melodic arch
  - Rodriguez-Zivic: perception & musical style
Corpus analysis

David Huron (1995):

The melodic arch in Western folksongs

corpus: 6251 folk songs from the ‘Essen Folksong Collection’
features: melodic pitch height, contour

Hypothesis: music theorists - melodic passages tend to exhibit an arch shape where the overall pitch contour rises and then falls over the course of a phrase or an entire melody

findings:
- tendency towards arch-shaped melodic contours confirmed
Corpus analysis

- in language studies: corpus linguistics
- in musicology:
  - statistical musicology
  - data-driven musicology
  - empirical musicology
  - ...
- examples:
  - Syncopation patterns in ragtime
  - Huron: the melodic arch
  - Rodriguez-Zivic: perception & musical style
Corpus analysis

Rodriguez-Zivic et al. (2011):
Perceptual basis of evolving Western musical styles
features: melodic pitch intervals, paired into ‘bigrams’ and clustered into 5 factors
findings:
  - baroque period music follows the diatonic scale closely (‘white keys on the piano’)
  - classical period works rely a lot on unison (repetition).
  - Romantic and post-romantic music expand these vocabulary of intervals
Corpus analysis

![Graph showing time series for Baroque, Classical, Romanticism, and XX Century with clusters 1, 2, 3, and 4.](image-url)
Corpus analysis

- dictionary based on pairs of melodic intervals used
- represents each 5-year period between 1730 and 1930 as a single, compact distribution
- $k = 5$ factors are then identified using $k$-means clustering
- four coincide with the historic periods of baroque, classical, romantic and post-romantic music
Corpus analysis

- Many more studies using symbolic data:
  - chords
    - De Clercq and Temperley (2011)
      - 99 rock songs, 20 for every decade 1950-2000
      - Analysis of chord root transitions and co-occurrence over time
      - Result: strong (but decreasing) prominence of the IV chord and the IV-I progression
    - Burgoyne (2013)
      - analysis of 1379 songs from Billboard dataset
      - Result: trend towards minor tonalities, decrease in the use of dominant chords, and a positive effect of ‘non-core’ roots (roots other than I, V, and IV) on popularity

- Today’s typology of corpus studies:
  - hypothesis-driven vs. discovery-driven
  - symbolic data vs. audio data
Audio features for corpus analysis

Main selection criteria for audio features:
- Features must have a clear natural language interpretation, so that results in the feature domain can be translated back into natural language.
- Features can only be used if they can be reliably computed.

Two example feature sets:
- Psycho-acoustic features
- Corpus-relative features

Psycho-acoustic features

signal measurements that correspond to human ratings of an attribute of sound
tested in a laboratory environment

- loudness
- sharpness
- roughness
Psycho-acoustic features

- loudness
- sharpness
- roughness

Equal-loudness contours (red) (from ISO 226:2003 revision)
Original ISO standard shown (blue) for 40-phon
Psycho-acoustic features

- loudness
  - intensity (in dB)
  - frequency content
- sharpness
- roughness

Equal-loudness contours (red) (from ISO 226:2003 revision)
Original ISO standard shown (blue) for 40-phon
Psycho-acoustic features

**Loudness**

- 1 sone = 1000 Hz at 40 dB (=40 phons)
- Sone is basis of ISO standard scale
- Sone is linear, phon logarithmic
Psycho-acoustic features

- loudness
- Sharpness
  - High frequency content
  - compute sharpness as weighted sum of the specific loudness levels in various bands
- roughness

Sharp:                             Unsharp:
Psycho-acoustic features

- loudness
- sharpness
- Roughness quantifies the subjective perception of rapid amplitude modulation of a sound

\[ R(X) = \sum_{f_i} \sum_{f_j} w(f_i, |f_j - f_i|) X(f_i) X(f_j) \]

rough

not rough
Psycho-acoustic features

- loudness
- sharpness
- Roughness: background *critical bandwidth*
  - filtering of frequencies within the cochlea
  - only if two frequency components are different enough, we perceive two different tones
  - if two frequency components are within the same critical bandwidth, we perceive them as one tone
  - Perceptual roughness of a complex sound (comprising many partials or pure tone components) depends on the distance between the partials measured in critical bandwidths.
  - A simultaneous pair of partials of about the same amplitude that is less than a critical bandwidth apart *produces roughness* associated with the inability of the basilar membrane to separate them clearly
Psycho-acoustic features

- loudness
- sharpness
- Roughness quantifies the subjective perception of rapid amplitude modulation of a sound

\[ R(X) = \sum_{f_i} \sum_{f_j} w(f_i, |f_j - f_i|) X(f_i) X(f_j) \]
Summary psycho-acoustic features

- Loudness
- sharpness
- roughness

- empirically established attributes of sound
- Attributes also used in natural language description of sound
Loudness and Dynamics

The ‘loudness war’:

1977

2013
Loudness and Dynamics

The ‘loudness war’: 

- Queen: “Seaside Rendez-vous” (1975)

Loudness and Dynamics

Deruty & Tardieu (2014):
Dynamic processing in mainstream music

corpus: 4500 tracks released between 1967 and 2011
(100 per year)

features: RMS, EBU-loudness, EBU-loudness range, peak-to-RMS factors
Dynamic processing in mainstream music

- RMS (root-mean square of the arithmetic mean)
  - Average loudness value during a certain time frame
- EBU Loudness
- EBU-loudness range
- Peak-to-RMS factors

RMS:
Dynamic processing in mainstream music

- RMS
- EBU Loudness (European Broadcasting Union)
- EBU-loudness range
- Peak-to-RMS factors

Loudness range:

“The difference between the 10th and 95th percentile of the distribution of 3 second loudness averages computed with 1 second overlap”

measures the variation of loudness on a macroscopic time-scale
Dynamic processing in mainstream music

- RMS
- EBU Loudness
- EBU-loudness range
- Peak-to-RMS factors

![Audio waveforms for Queen and Red Hot Chili Peppers](image)
Dynamic processing in mainstream music

- RMS
- EBU Loudness
- EBU-loudness range
- Peak-to-RMS factors (measures micro dynamics)
Loudness and Dynamics

Deruty & Tardieu (2014): Dynamic processing in mainstream music corpus: 4500 tracks released between 1967 and 2011 (100 per year) features: RMS, EBU-loudness, EBU-loudness range, peak-to-RMS findings:
Loudness and Dynamics

Deruty & Tardieu (2014):
Dynamic processing in mainstream music
corpus: 4500 tracks released between 1967 and 2011
(100 per year)
features: RMS, EBU-loudness, EBU-loudness range
findings:
- Loudness and RMS increase, with a peak around 2007
- Micro-dynamics have decreased as loudness went up
- Macro-dynamics (loudness range) have not decreased
Application of psycho-acoustic features to chorus analysis
Chorus analysis

  ±7000 song sections, 1958 - 1992
features: loudness, loudness range, sharpness, roughness
  + a few others re: pitch height and timbre variance

What makes a chorus distinct from other sections in a song?
Why chorus analysis?

• Choruses: more prominent, more catchy, more memorable than other sections in a song

• MIR: chorus detection primarily based on identifying the most-repeated section in a song.

• chorus detection is tied to audio thumbnailing, music summarization, structural segmentation

• Question: Can we use computational methods to improve our understanding of choruses?
Chorus analysis

analysis method:
learning a probabilistic graphical model: (based on 11 perceptual features and chorusness variables)
Chorus analysis

Van Balen, Burgoyne, Wiering, Veltkamp (2013):
An analysis of chorus features in popular song corpus: Billboard dataset
±7000 song sections, 1958 - 1992
features: loudness, loudness range, sharpness, roughness
+ a few others re: pitch height and timbre variance
findings:

- MFFC variance
- Roughness
- Loudness
- Pitch Centroid
- Pitch Salience
- Sharpness
- Loudness range

more diverse timbre
 timbre more rough
 choruses are louder
 higher pitch
 more salient pitch
 more high frequencies
 less dynamics
Corpus analysis: Where to look for the hook

A study of catchiness in popular songs

- what parts of songs are easily remembered?
  what is the hook?

- how important is repetition
  striking moment vs. recurring riff

- what role does expectation play?
  surprise vs. cliché
Where to look for the hook

a study of *catchiness* in popular songs

- what parts of songs are easily remembered?
  - what is the *hook*?

- how important is repetition
  - striking moment vs. recurring riff

- what role does expectation play?
  - surprise vs. *cliché*
Where to look for the hook

a study of catchiness in popular songs

- what parts of songs are easily remembered?
  what is the hook?
Where to look for the hook

Hooked!
a game-with-a-purpose to study catchiness

- Players get 15 s to recognize a song.
- If yes, the song mutes for 4 seconds.
- When it comes back, does it come back in the right place?
Where to look for the hook

a study of catchiness in popular songs

- what parts of songs are easily remembered?
  what is the hook?

- how important is repetition
  striking moment vs. recurring riff

- what role does expectation play?
  surprise vs. cliché
Hook analysis

Van Balen, Burgoyne, Bountouridis, Müllensiefen, Veltkamp (2015):

Corpus Analysis Tools for Computational Hook Discovery

corpus: Hooked! data

1750 song segments from 321 songs and 973 players
features: chorus features + melody and harmony features
   + corpus-relative features based on the above
Where to look for the hook

Corpus-relative features

- Second order features
Where to look for the hook

Corpus-relative features

- Second order features
  - Symbolic (e.g. FANTASTIC toolbox): discrete numbers (countable)
Where to look for the hook

Corpus-relative features

- Second order features
  - Symbolic (e.g. FANTASTIC toolbox): discrete numbers (countable)
  - Audio: continuous, uninterrupted signals
    - Features measured over short windows, represent continuous, uncountable quantities
Where to look for the hook

Corpus-relative features

- Features in their raw form are not always informative
- Therefore: convert a features to a scale of common vs. uncommon.

For 1-dimensional features (e.g. loudness): \( f(X) \) probability density estimate

\[
Z(X) = \logit \left[ \frac{\text{rank} (f(X))}{N} \right]
\]

i.e., a non-parametric scaling of a feature values frequency
Where to look for the hook

Corpus-relative features

- Features in their raw form are not always informative
- Therefore: convert a features to a scale of common vs. uncommon.

For 1-dimensional features (e.g. loudness): \( f(X) \) probability density estimate

\[
Z(X) = \text{logit} \left[ \frac{\text{rank} \left( f(X) \right)}{N} \right]
\]

\( N \): size of reference corpus

i.e., a non-parametric scaling of a feature values frequency
Where to look for the hook

Corpus-relative features

- Features in their raw form are not always informative
- Therefore: convert a features to a scale of common vs. uncommon.

For n-dimensional features (e.g. pitch class distribution):

\[ I(F) = - \sum_{i=1}^{d} F(i) \log F_c(i) \]

i.e., information: how much information does an observed distribution provide compared to a corpus average (measure of unexpectedness)
Where to look for the hook

Corpus-relative features

- Features in their raw form are not always informative
- Therefore: convert features to a scale of common vs. uncommon
- Reference corpus can be varied:
  - large corpus as reference
    → feature measures conventionality
  - sections from the same song as reference
    → feature measures recurrence
Hook analysis


corpus: Hooked! data

1750 song segments from 321 songs and 973 players
features: chorus features + melody and harmony features

+ corpus-relative features based on the above findings:
Hook analysis

findings:
8 components correlate significantly

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Audio&lt;sup&gt;a&lt;/sup&gt;</th>
<th></th>
<th>Audio&lt;sup&gt;b&lt;/sup&gt;</th>
<th></th>
<th>Symbolic&lt;sup&gt;b&lt;/sup&gt;</th>
<th></th>
<th>Combined&lt;sup&gt;b&lt;/sup&gt;</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\beta} )</td>
<td>99.5 % CI</td>
<td>( \hat{\beta} )</td>
<td>99.5 % CI</td>
<td>( \hat{\beta} )</td>
<td>99.5 % CI</td>
<td>( \hat{\beta} )</td>
<td>99.5 % CI</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.84 [-0.91, -0.77]</td>
<td></td>
<td>-0.67 [-0.78, -0.56]</td>
<td></td>
<td>-0.62 [-0.73, -0.51]</td>
<td></td>
<td>-0.63 [-0.74, -0.53]</td>
<td></td>
</tr>
<tr>
<td>Audio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocal Prominence</td>
<td>0.14 [0.10, 0.18]</td>
<td></td>
<td>0.11 [0.04, 0.17]</td>
<td></td>
<td>0.08 [0.01, 0.15]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timbral Conventionality</td>
<td>0.09 [0.05, 0.13]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melodic Conventionality</td>
<td>0.06 [0.02, 0.11]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M/H Entropy Conventionality</td>
<td>0.06 [0.02, 0.10]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharpness Conventionality</td>
<td>0.05 [0.02, 0.09]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harmonic Conventionality</td>
<td>0.05 [0.01, 0.10]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timbral Recurrence</td>
<td>0.05 [0.02, 0.08]</td>
<td></td>
<td>0.07 [0.02, 0.13]</td>
<td></td>
<td>0.07 [0.01, 0.12]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mel. Range Conventionality</td>
<td>0.05 [0.01, 0.08]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symbolic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melodic Repetitivity</td>
<td></td>
<td></td>
<td>0.12 [0.06, 0.19]</td>
<td></td>
<td>0.11 [0.05, 0.17]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mel/Bass Conventionality</td>
<td></td>
<td></td>
<td>0.07 [0.01, 0.13]</td>
<td></td>
<td>0.08 [0.01, 0.14]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hook analysis

Van Balen, Burgoyne, Müllensiefen, Veltkamp (in review): Corpus Analysis Tools for Hook Discovery

corpus: Hooked! data

1750 song segments from 321 songs and 973 players

features: chorus features + melody and harmony features + corpus-relative features based on the above findings:

- features correlated with vocals predict hooks best
- conventionality dominates the remainder of the results
- recurrence also contributes
Conclusions

- quality of corpus studies also depends on choice of data and analysis method, but generally...
- good features have a clear natural language interpretation, so that results in the feature domain can be translated back into natural language
- ...and can be reliably computed

Two types of feature that address these criteria:
- psycho-acoustic features
- corpus-relative features
Summary

- Use of audio features for characterizing corpora
- Features for characterizing evolution
- Very important for classification of styles
- Games and catchy music
References


