Audio Feature Extraction for Corpus Analysis

Anja Volk
Sound and Music Technology

17 Dec 2019

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: audio feature extraction

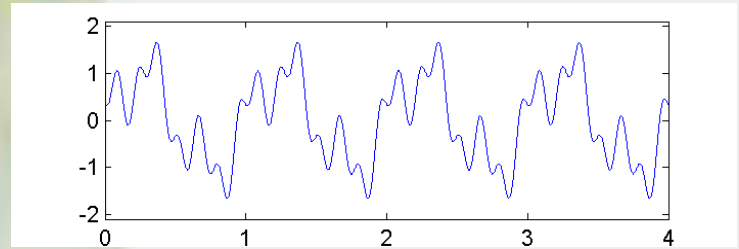
- Fourier analysis: what is it for? What kind of information do we get out of it?
  - Analyzing the frequency content of a signal

Recap: audio feature extraction

Idea: Decompose a given signal into a superposition of sinusoids (elementary signals).

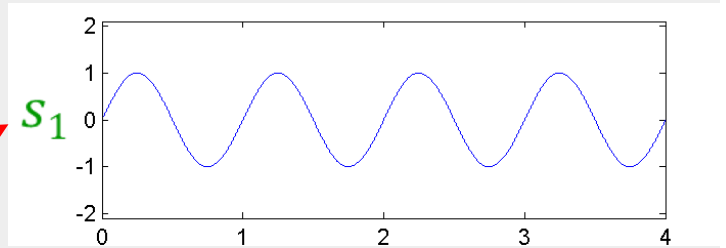
$$f = s_1 + s_2 + s_3$$

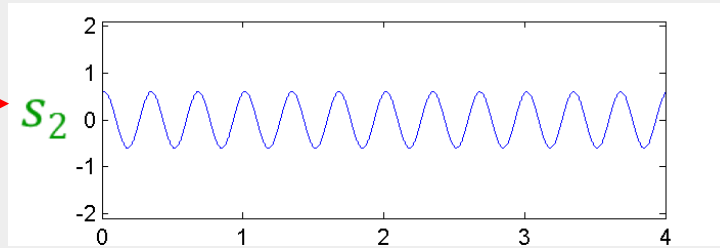
Signal f

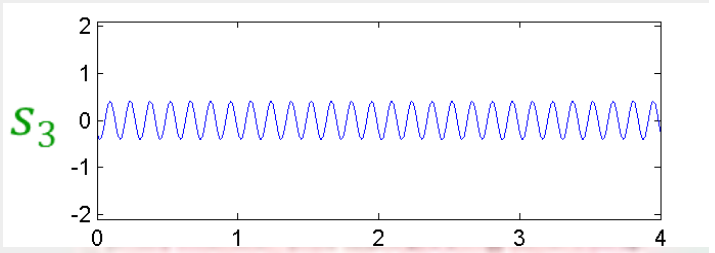


Time (seconds)

Sinusoids







Time (seconds)

Recap: audio feature extraction

- Fourier analysis: what is it for?
  - Analyzing the frequency content of a signal
- What is the use of STFT?
  - Provides analysis over time

Short Term Fourier Transformation

- Problem Fourier:
  - Tells **which** frequencies occur, but does not tell **when** the frequencies occur.
  - Frequency information is averaged over the entire time interval.

Short Term Fourier Transformation

- Problem Fourier:
  - Tells which frequencies occur, but does not tell when the frequencies occur.
  - Frequency information is averaged over the entire time interval.
- STMT: Consider only a small section of the signal or the spectral analysis
  - Section is determined by pointwise multiplication of the signal with a localizing window function of the signal with a localizing window function

Recap: audio feature extraction

- Fourier analysis: what is it for?
  - Analyzing the frequency content of a signal
- What is the use of STFT?
  - Provides analysis over time
  - Heisenberg Uncertainty
    - Size of window constitutes a trade-off between time resolution and frequency resolution:
      - Large window results in poor time resolution but good frequency resolution;
      - Small window : good time resolution but poor frequency resolution
- How are chroma features derived?
  - Sum the spectral energy per octave;
  - Chroma features are suitable for chord and key detection
- Beat tracking (Dan Ellis, 2007)

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

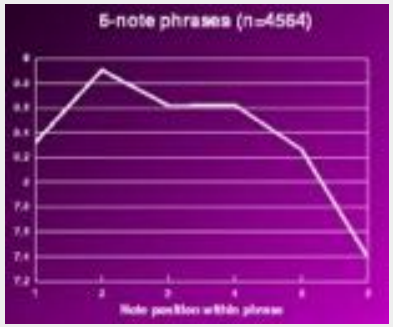
corpus: 6251 folk songs from the 'Essen Folksong Collection'

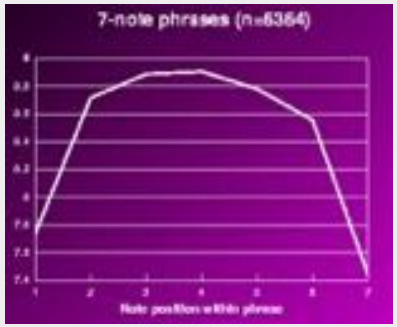
features: melodic pitch height, contour

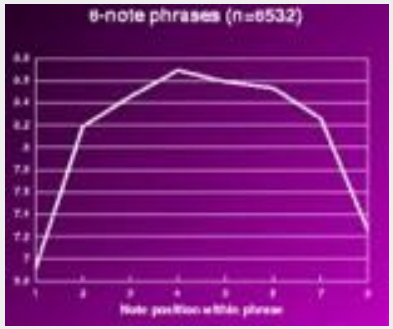
Hypothesis: music theorists - melodic passages tend to exhibit an *arch* shape where the overall pitch contour rises and then falls over the course of a phrase or an entire melody

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

corpus: 'Peachnote' corpus of classical music, http://www.peachnote.com/info.html

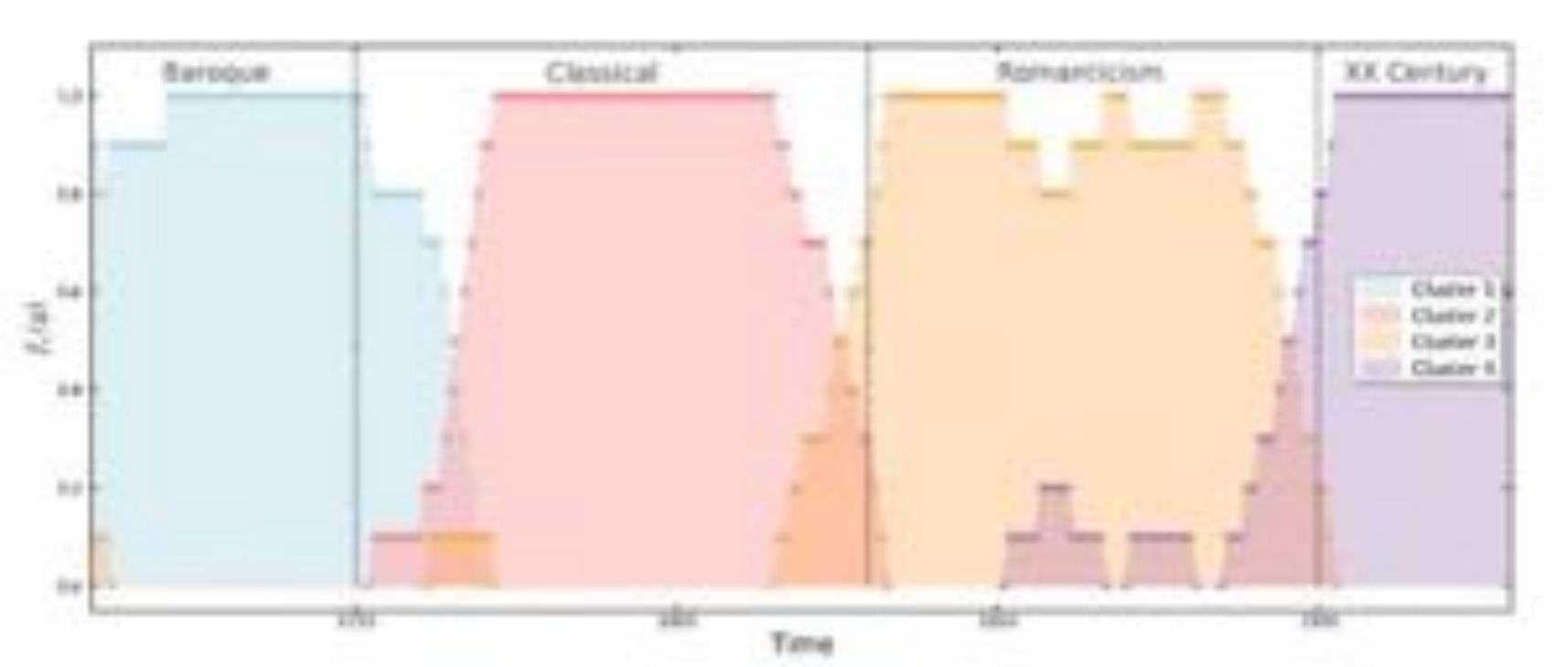
features: melodic pitch intervals, paired into 'bigrams' and clustered into 5 factors

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

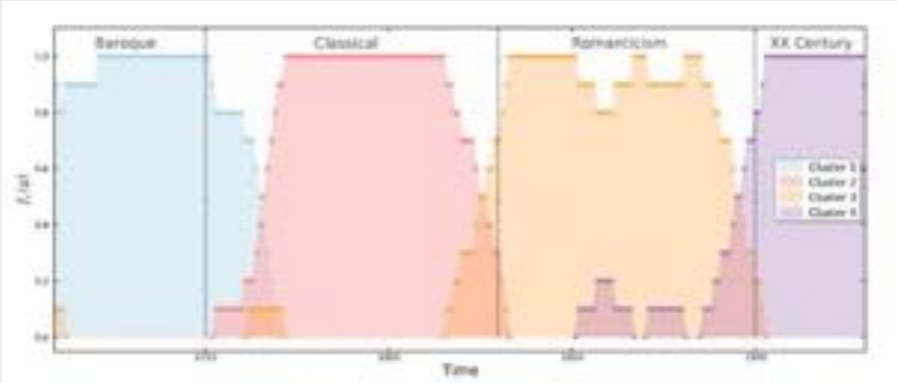



Corpus analysis



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
- Baroque: use of the diatonic scale, Classic: repeated notes, Romanticism: wide harmonic intervals, post-modern: chromatic tonality.



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
  - rhythmic motives: Mauch et al (2012), Volk & De Haas (2013)
- 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

- PhD Thesis Jan van Balen: Audio Description and Corpus Analysis of Popular Music, 2016, Utrecht University

Psycho-acoustic features

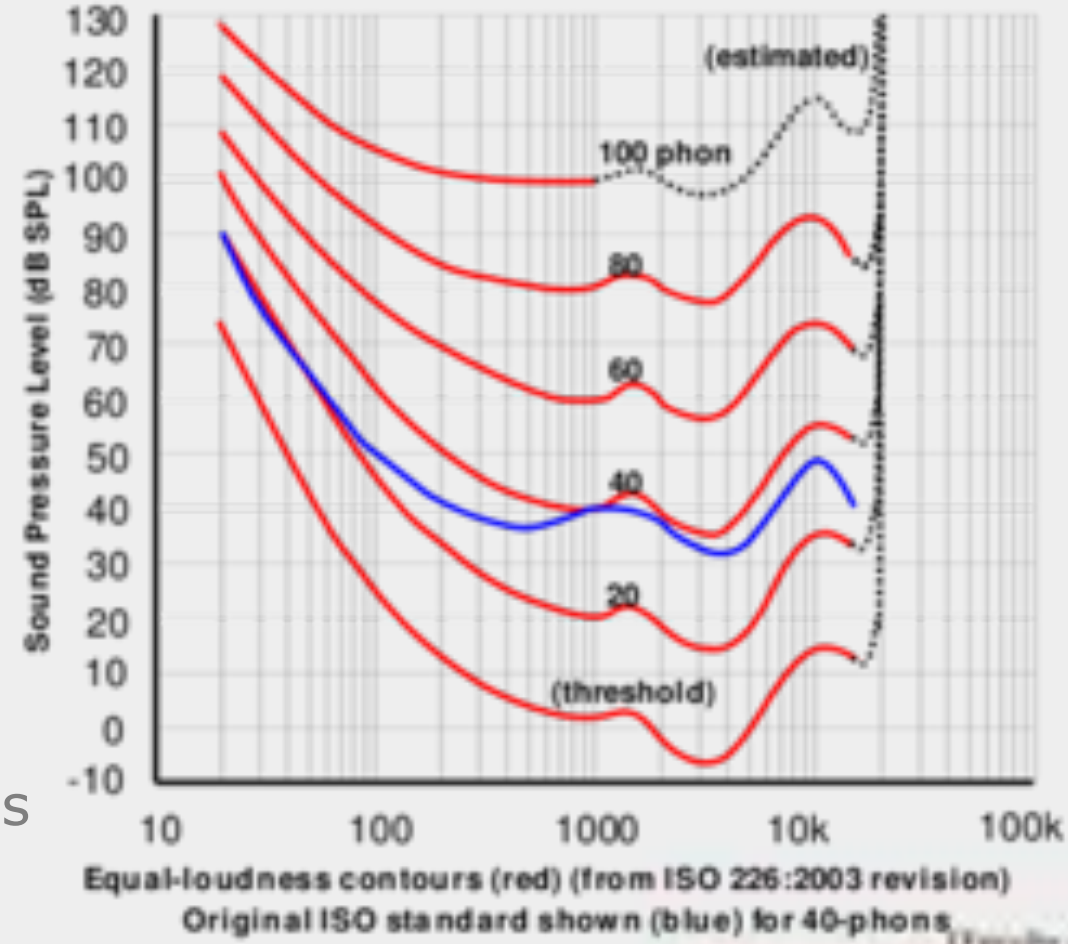
signal measurements that correspond to human ratings of an attribute of sound

tested in a laboratory environment

- loudness
- sharpness
- roughness

Psycho-acoustic features

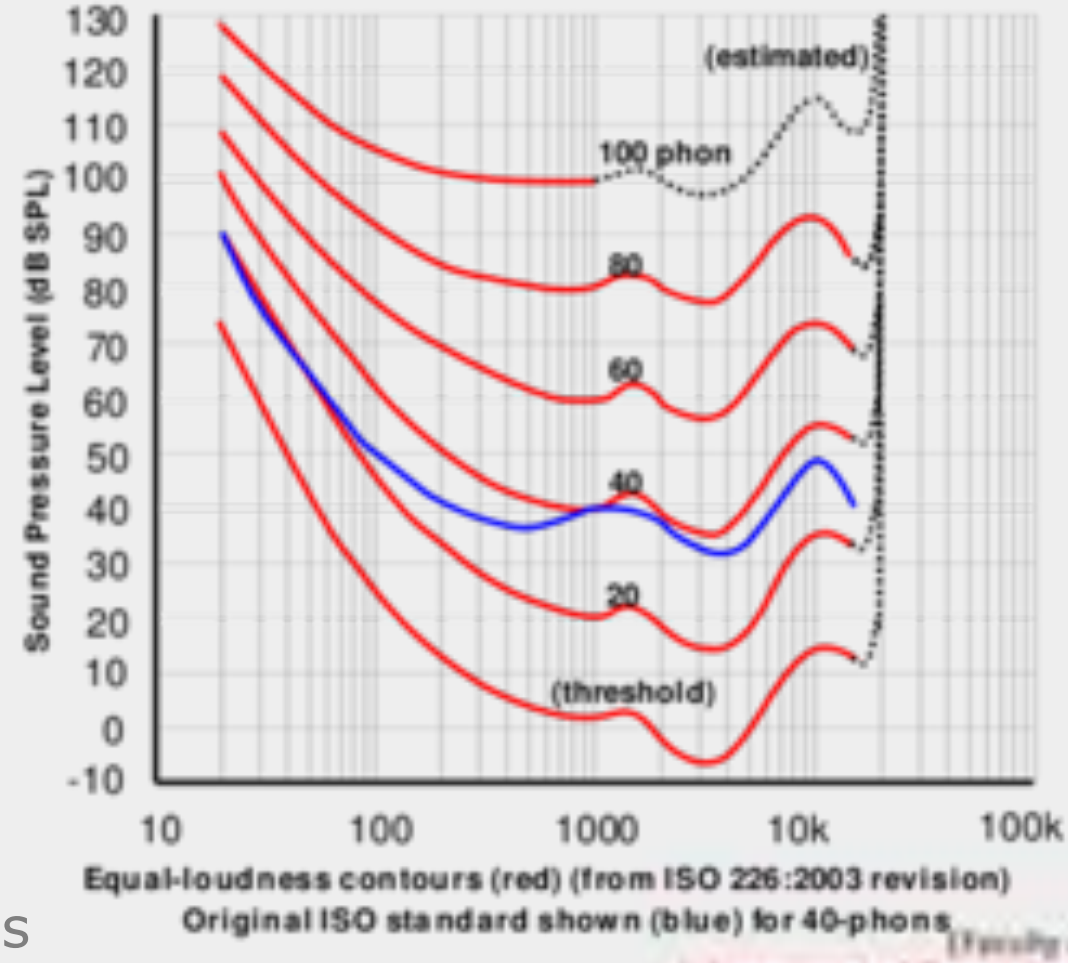
- loudness
- sharpness
- roughness



wikimedia commons

Psycho-acoustic features

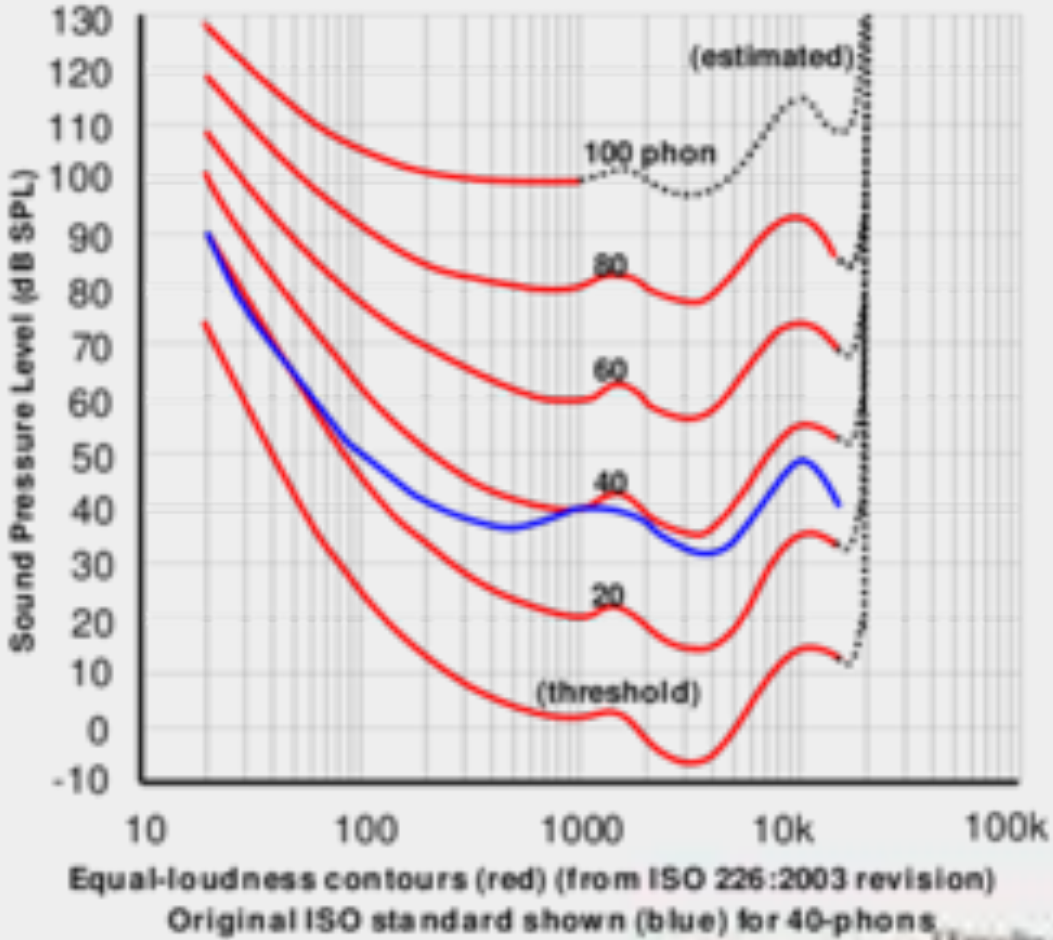
- loudness
  - intensity (in dB)
  - frequency content
- sharpness
- roughness



wikimedia commons

Psycho-acoustic features

- loudness
- sone and phone
- 1 sone = 1000 Hz at 40 dB (=40 phons)
- Sone is basis of ISO standard scale
- sone is linear, phon logarithmic



wikimedia commons

Psycho-acoustic features

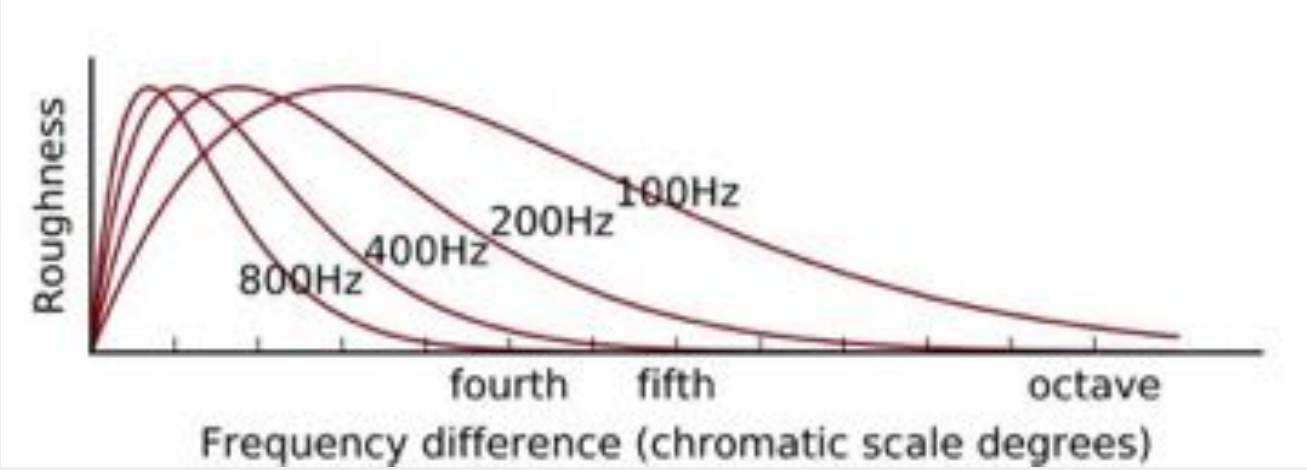
- loudness
- Sharpness
  - High frequency content
  - compute sharpness as weighted sum of the specific loudness levels in various bands
- roughness

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

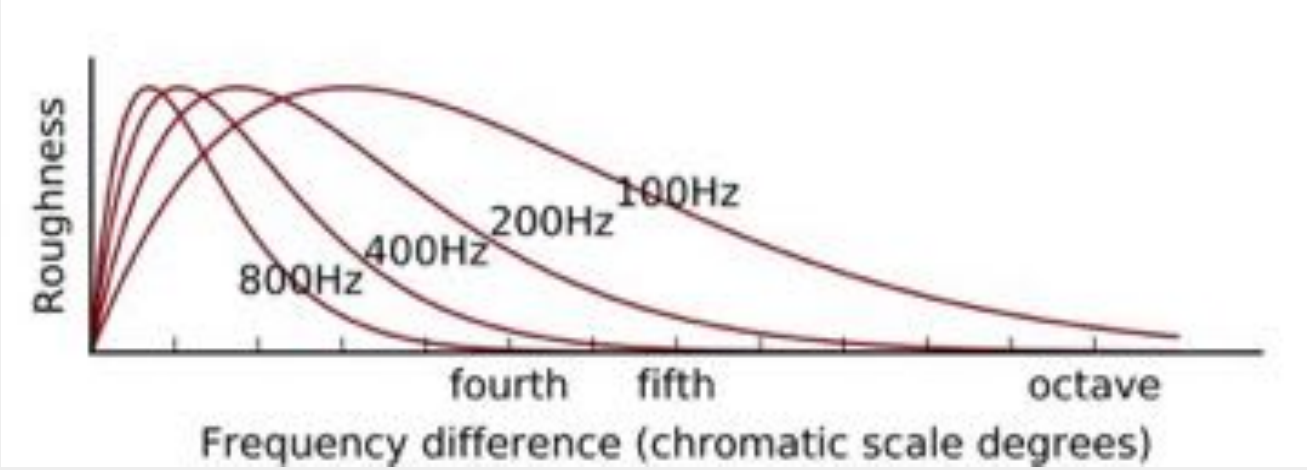
not rough

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



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

rough

not rough

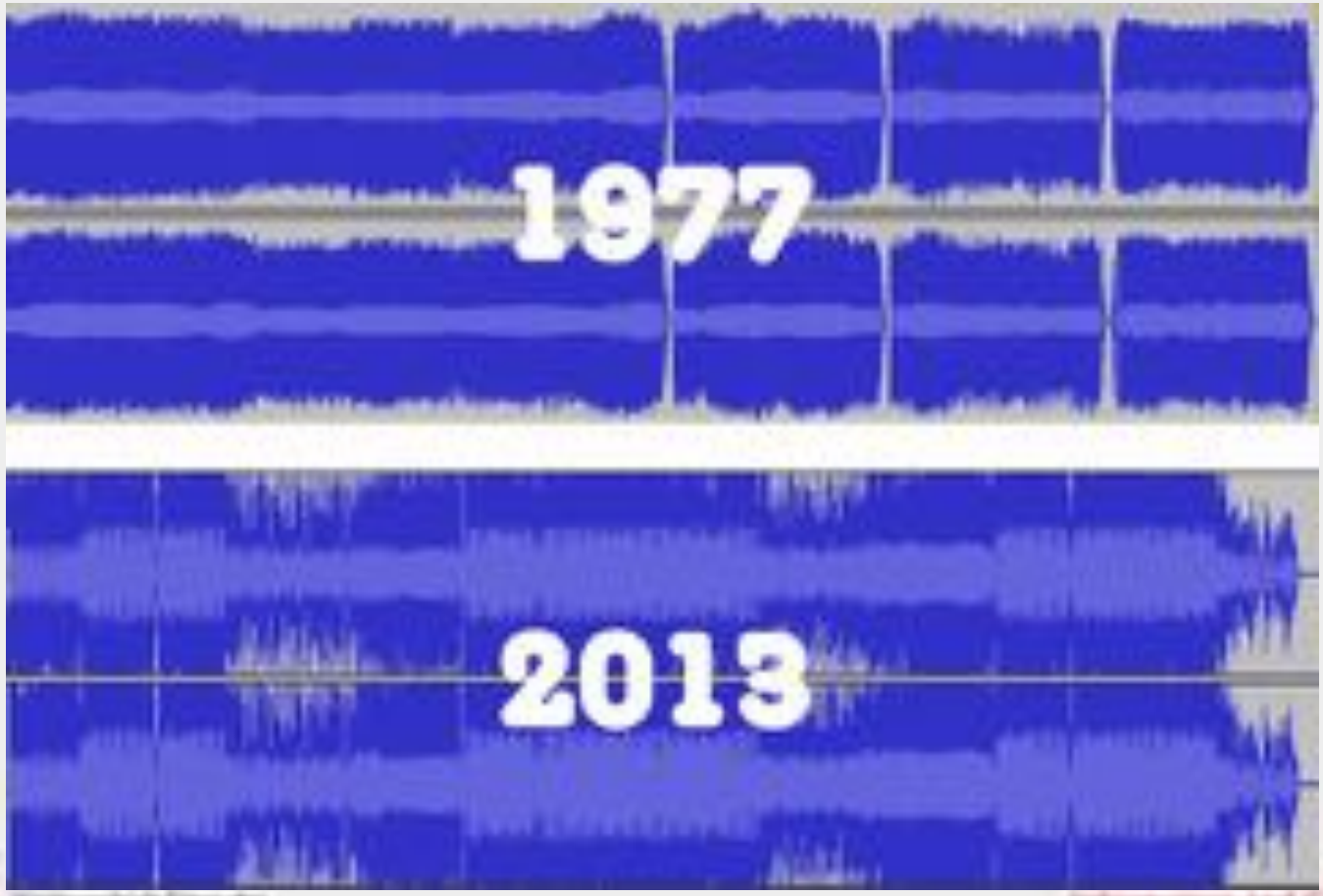
Summary psycho-acoustic features

- Loudness
- sharpness
- roughness

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

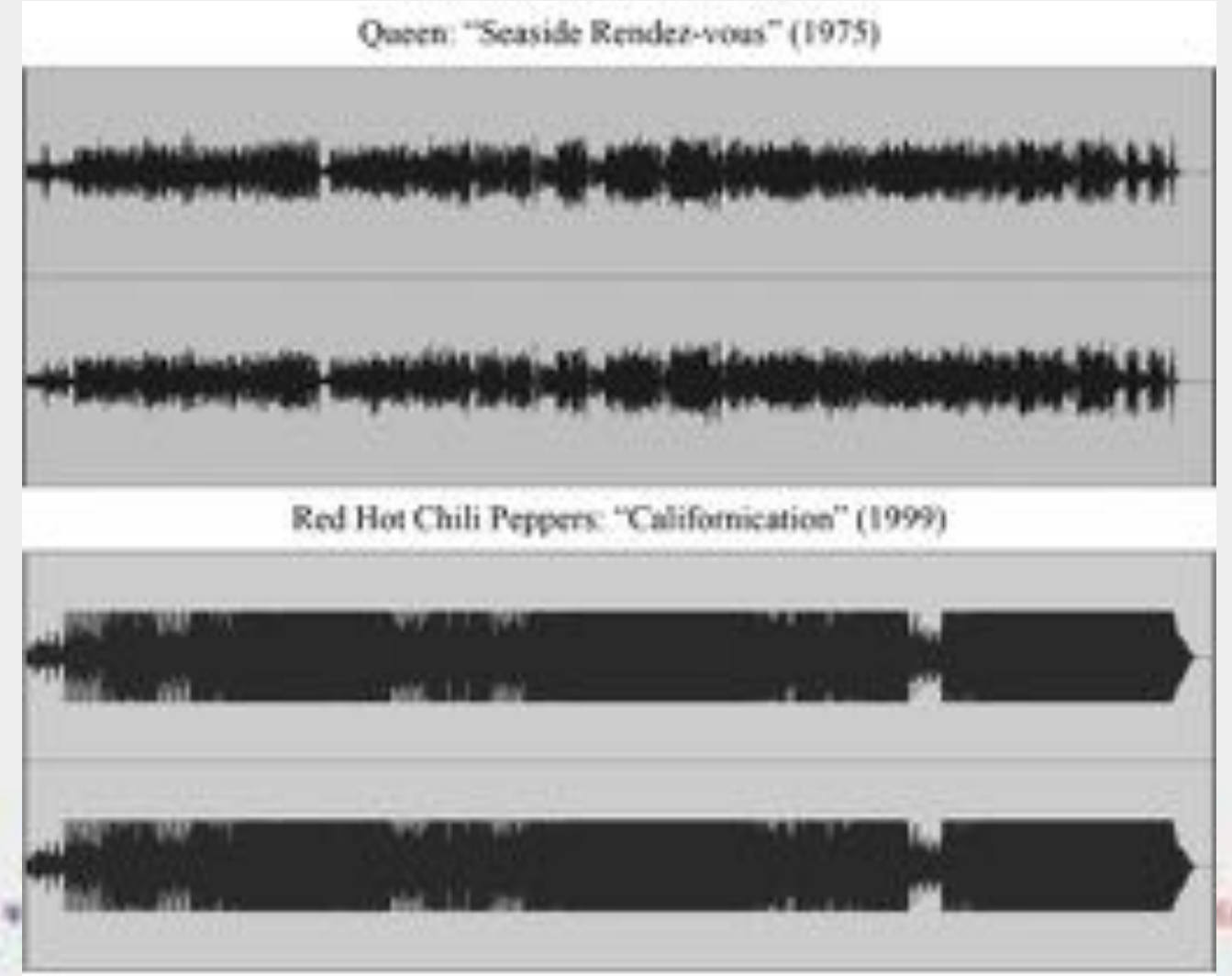
Loudness and Dynamics

The ‘loudness war’:



Loudness and Dynamics

The ‘loudness war’:



Loudness and Dynamics

Deruty & Tardieu (2014):

Dynamic processing in mainstream music

corpus: 4500 tracks released between 1967 and 2011

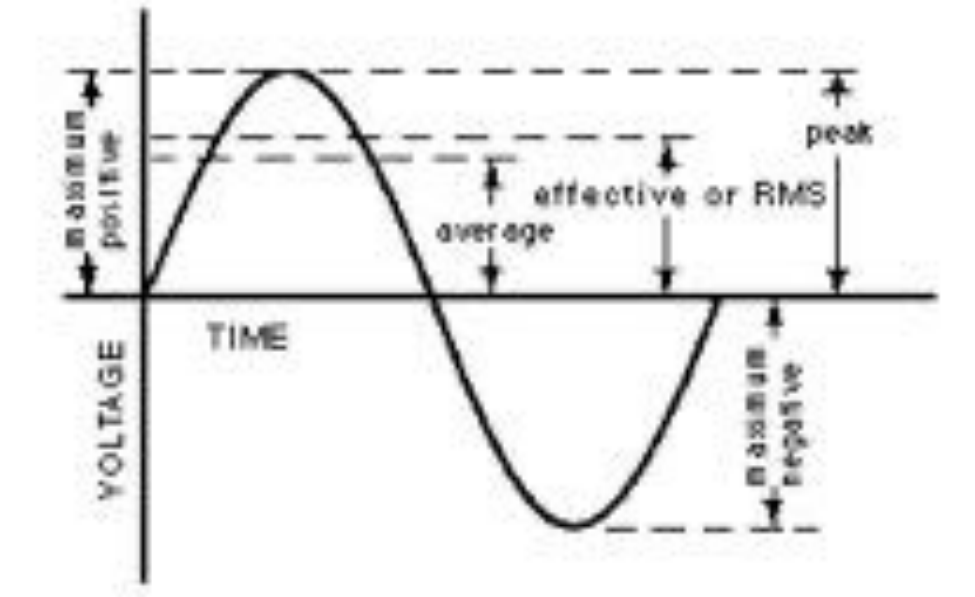
(100 per year)

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
  - Not the momentary peak level, but average
- EBU Loudness
- EBU-loudness range
- Peak-to-RMS factors

RMS:



Dynamic processing in mainstream music

- RMS
- EBU Loudness (European Broadcasting Union)
- EBU-loudness range
  - measures the variation of loudness on a macroscopic time-scale, in units of LU (Loudness Units).
  - used to measure the long term variability of loudness, high values showing that there are some very quiet parts and some very loud parts in a track
  - suitable feature for the evaluation of musical dynamics in the classical sense, such as *pianissimo* to *fortissimo*.
- Peak-to-RMS factors

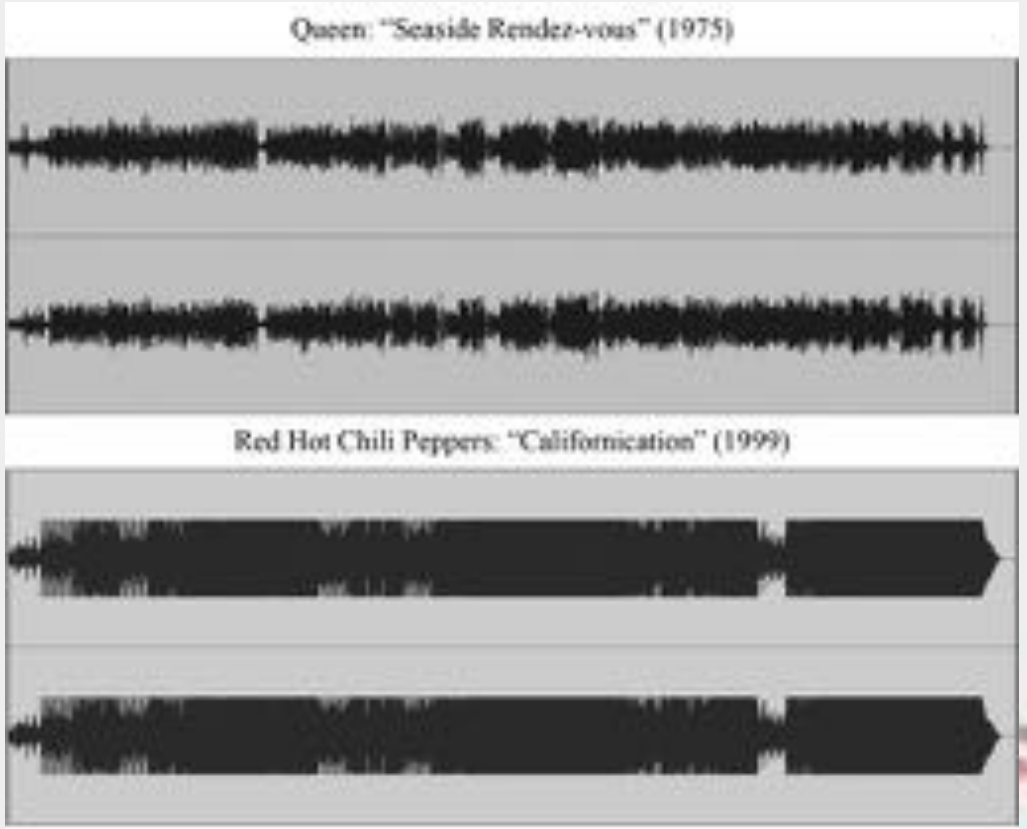
Loudness range:

"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

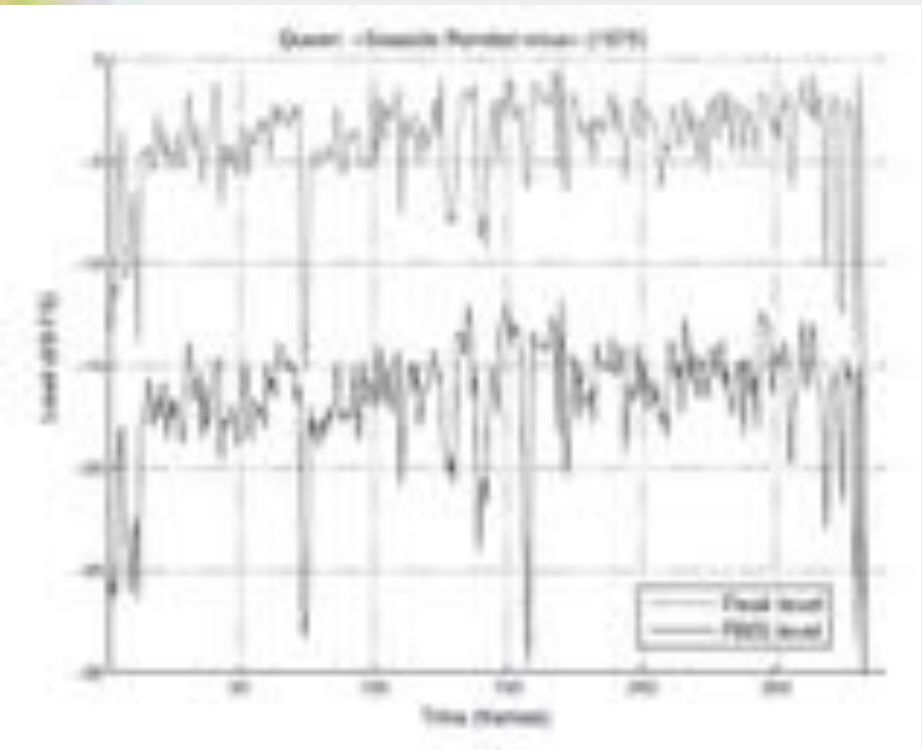
Dynamic processing in mainstream music

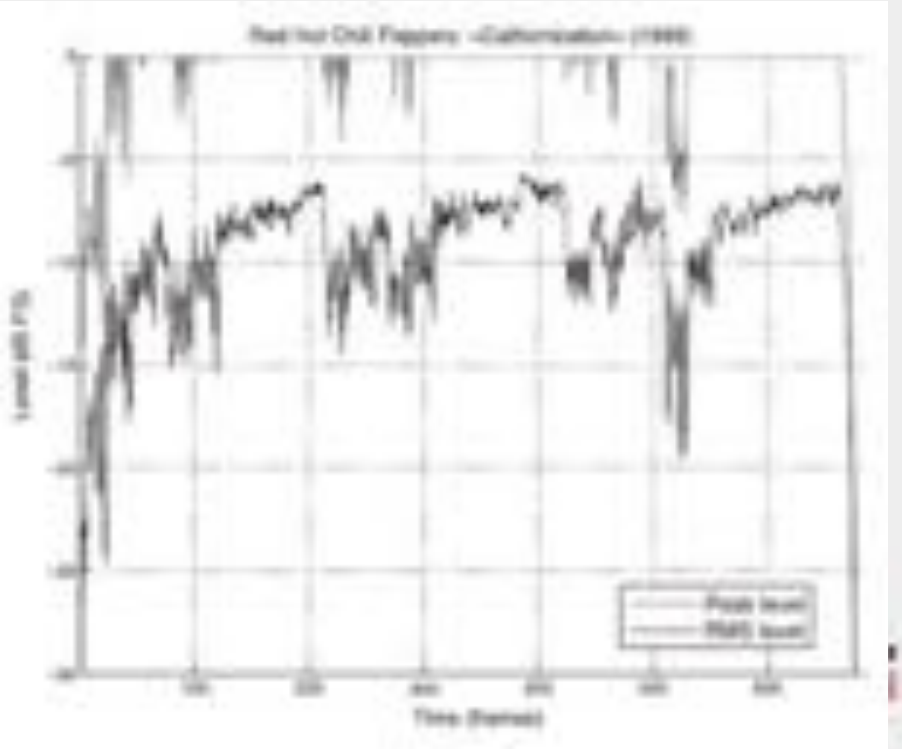
- RMS
- EBU Loudness
- EBU-loudness range
- Peak-to-RMS factors



Dynamic processing in mainstream music

- RMS
- EBU Loudness
- EBU-loudness range
- Peak-to-RMS factors (measures micro dynamics)
  - Factor close to 1: RMS and peak levels evolve together
  - Factor close to 0: peak levels stable regardless of RMS level





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

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

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	more diverse timbre
Roughness	timbre more rough
Loudness	choruses are louder
Pitch Centroid	higher pitch
Pitch Salience	more salient pitch
Sharpness	more high frequencies
Loudness range	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

- Features in their raw form are not always informative
- Therefore: convert a 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

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

Hook analysis

findings:

8 components correlate significantly

Parameter	Audio[a]		Audio[b]		Symbolic[b]		Combined[b]	
	$\hat{\beta}$	99.5% CI	$\hat{\beta}$	99.5% CI	$\hat{\beta}$	99.5% CI	$\hat{\beta}$	99.5% CI
				Fixed effects				
Intercept	−0.84	[−0.91, −0.77]	−0.67	[−0.78, −0.56]	−0.62	[−0.73, −0.51]	−0.65	[−0.74, −0.53]
Audio								
Vocal Prominence	0.14	[0.10, 0.18]	0.11	[0.04, 0.17]			0.08	[0.01, 0.15]
Timbral Conventionality	0.09	[0.05, 0.13]						
Melodic Conventionality	0.06	[0.02, 0.11]						
MSI Entropy Conventionality	0.06	[0.02, 0.10]						
Sharpness Conventionality	0.05	[0.02, 0.09]						
Harmonic Conventionality	0.05	[0.01, 0.10]						
Timbral Recurrence	0.05	[0.02, 0.08]						
Met. Range Conventionality	0.05	[0.01, 0.08]	0.07	[0.02, 0.13]			0.07	[0.01, 0.12]
Symbolic								
Melodic Repetitivity					0.12	[0.06, 0.19]	0.11	[0.05, 0.17]
Met./Bass Conventionality					0.07	[0.01, 0.15]	0.08	[0.01, 0.14]

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

- David Huron (1995). The melodic arch in Western folksongs. *Computing in Musicology*, Vol. 10, pp. 3-23.
- John Ashley Burgoyne, Jonathan Wild, and Ichiro Fujinaga. Compositional Data Analysis of Harmonic Structures in Popular Music. Mathematics and Computation in Music, pages 52–63, 2013.
- Trevor de Clercq and David Temperley. A corpus analysis of rock harmony. Popular Music, 30(01):47–70, jan 2011.
- Rodriguez-Zivic, Shifres & Cecchi (2011). Perceptual basis of evolving Western Musical Styles. *Proceedings of the National Academy of Science,* Vol. 110, pp. 10034-10038,
- Deruty & Tardieu (2014). Dynamic processing in mainstream music. *Journal of the Audio Engineering Society*, Volume 62, pp. 42-55,
- Van Balen, Burgoyne, Wiering, Veltkamp (2013). An analysis of chorus features in popular song. *Proceedings of the 14th Society of Music Information Retrieval Conference (ISMIR).*
- Van Balen, Burgoyne, Bountouridis, Müllensiefen, Veltkamp (2015). Corpus Analysis Tools for Computational Hook Discovery. ISMIR proceedings