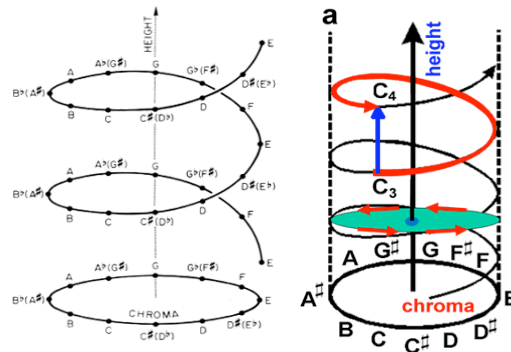# Characterizing harmony from audio

---

Pitch Helix

- The pitch helix is a pitch space where linear pitch is wrapped around a cylinder, thus modeling the special relationship that exists between octave intervals


- The model is a function of 2-dimensions:
- Height: naturally organizes pitches from low to high
- Chroma: represents the inherent circularity of pitch

Shepard Tones

- Sound made out of a superposition of octave-related sinusoids
- Uses only even partials weighted by a bell-shaped envelope
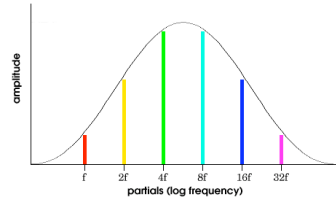
- A Shepard scale is created when the f0 of the sound is progressively changed, thus creating the illusion of constantly rising/falling tones
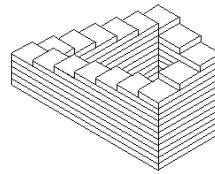

- Shepard/Tenney and Risset's continuous glissandi

Chroma: Pitch Class Profile

- Chroma describes the angle of the pitch rotation as it travels the helix
- Two octave-related pitches will share the same angle in the chroma circle.
- In the western tonal scale this angle can only take one of 12 possible values or pitch classes, thus the chroma can be seen as a pitch class profile.
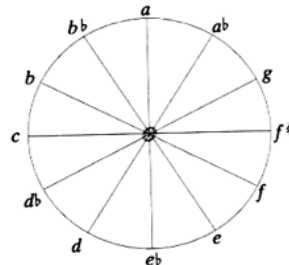

- A chord can be described as a function of its pitch classes
- Chroma representation is usually considered to be well suited for modelling harmony

Calculating chromas (1)

a) Calculate the FFT of a signal segment

b) Each FFT bin is mapped to its closest note, according to:

$$f(p) = 440 \cdot 2^{(p-69)/12}$$

where p is the note number. This is equivalent to segmenting the spectrum into note regions (± 1/4 tone from note center)
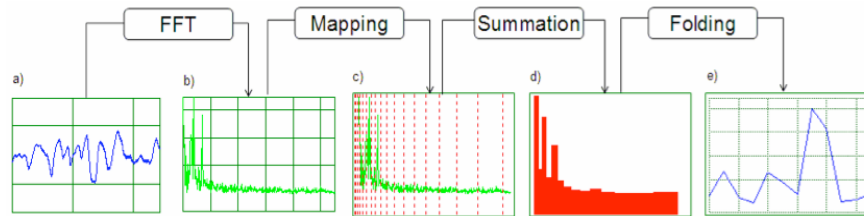

c) The average amplitude within regions is calculated

d) Resulting histogram is folded, collapsing bins belonging to the same pitch class into one.

Calculating chromas (2)

- This process is equivalent to using a Constant-Q filterbank where the filters' center frequencies are defined as:

$$f_k = 2^{k/\beta} f_{\min}$$

- With f_{min} as the minimum (or reference) frequency, and β as the analysis resolution (number of bins per octave)

- The bandwidth (BW) of each filter changes to maintain the f_k/BW ratio (Q) constant.

- The shape of the filter frequency response is important to weight according to the distance to the note's frequency
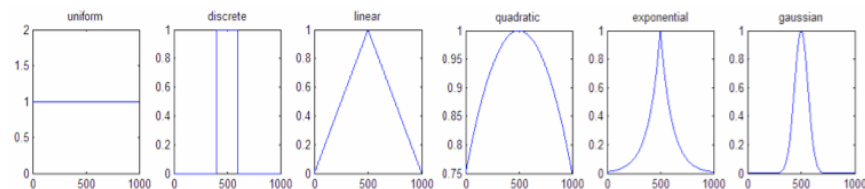
Calculating chromas (3)

- Linear vs Logarithmic Spectral Analysis for Music:
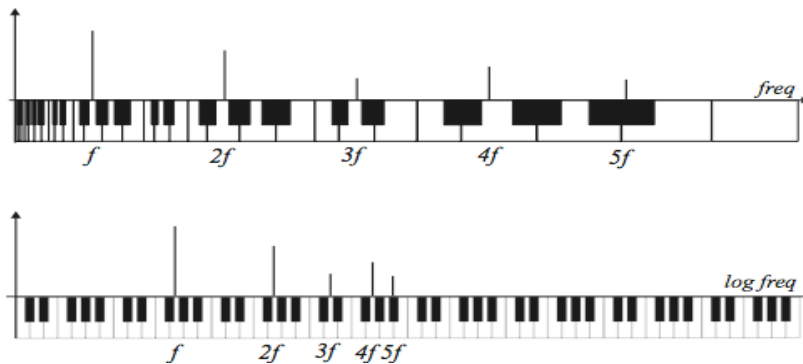
Calculating chromas (4)

- The previous approach to chroma calculation relies on the linear resolution of the FFT for its information gathering
- An alternative to this is the Constant-Q transform

$$X_{cq}(k) = \frac{1}{N(k)} \sum_{n=0}^{N(k)-1} x(n)w(n,k)e^{-j2\pi Qn/N(K)}$$
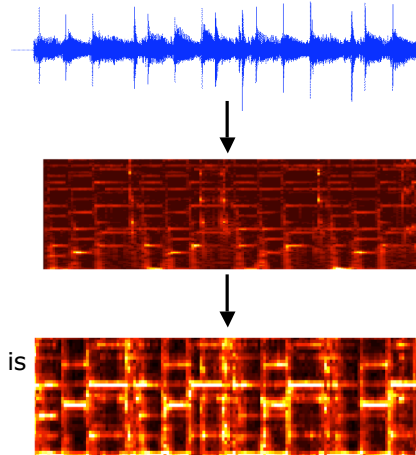
$$N(k) = f_s Q / f_k$$

- That uses a variable window length to obtain more resolution at lower frequencies and less at higher (logarithmic distribution of bins in frequency)

Calculating chromas (5)

- X_{cq} -> Constant-Q transform
- M -> total numbers of octaves
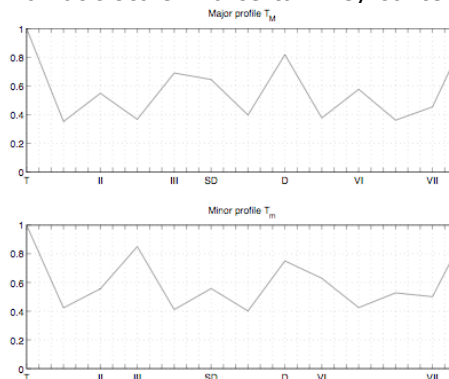- X_{cq} can be fold into a chroma:

$$Chroma(b) = \sum_{m=0}^{M} \left| X_{cq}(b + m\beta) \right|$$

- $b\ in\ [1,\beta]$ -> chroma bin number
- The sequence of chroma vectors is known as a *chromagram*
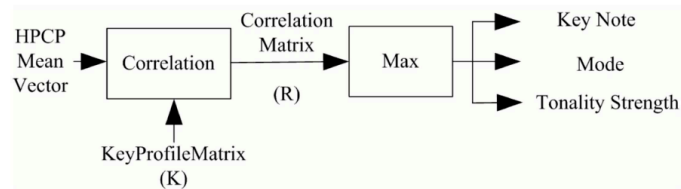

Estimating key (1)

- During the 1980s Krumhansl and her colleagues performed a number of subjective studies measuring the expectation of each tone in the chromatic scale in a certain key context

- As a result they proposed a probe tone model of key profiles characterizing tone likelihoods for major and minor keys
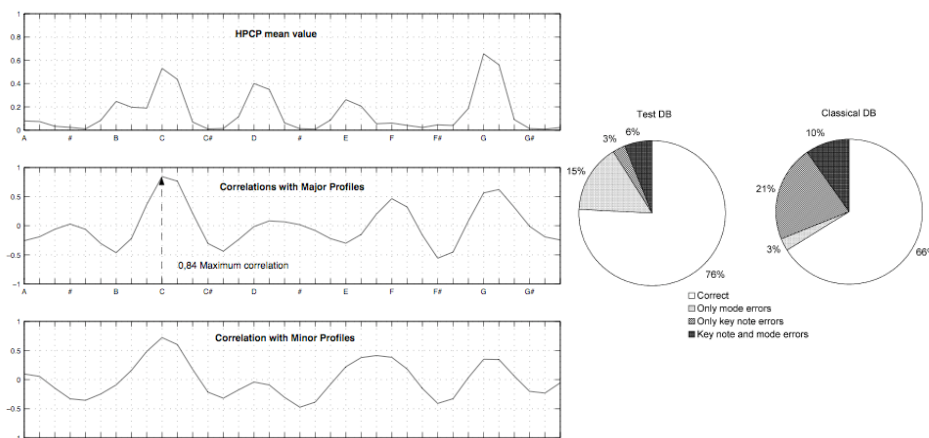
Estimating key (2)

- A number of researchers have used (versions of) those profiles to estimate the key of a musical piece, e.g. Gomez and Herrera (2004), Pauws (2004)

- The idea is that there is strong cross-correlation between the information on those profiles and the chromas (for segment key) or mean chromas (for entire songs) extracted from the signal
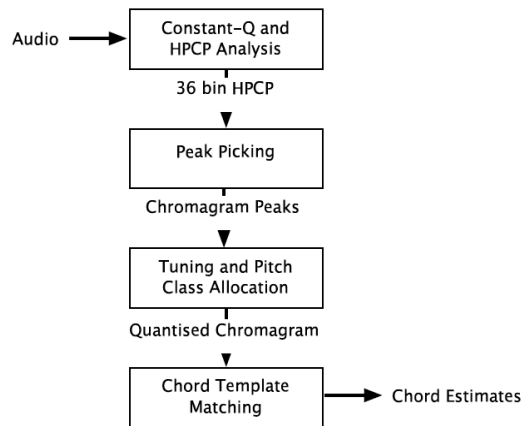
Estimating key (3)

Gomez and Herrera, ISMIR 2004

Chord Estimation (1)

- Chords can be estimated by post-processing the chroma and matching with simple chord templates, e.g. Harte and Sandler (2005)

Audio → Constant−Q and HPCP Analysis

36 bin HPCP

↓

Peak Picking

Chromagram Peaks

↓

Tuning and Pitch Class Allocation

Quantised Chromagram

↓

Chord Template Matching → Chord Estimates

Chord Estimation (2)

- Chromagram tuning
- β = 36 bins per octave -> 3 bins per note

Tuning deviation

Peak picking

Position of Peaks

Time Frames

flat ← 0 → sharp

Chord Estimation (3)
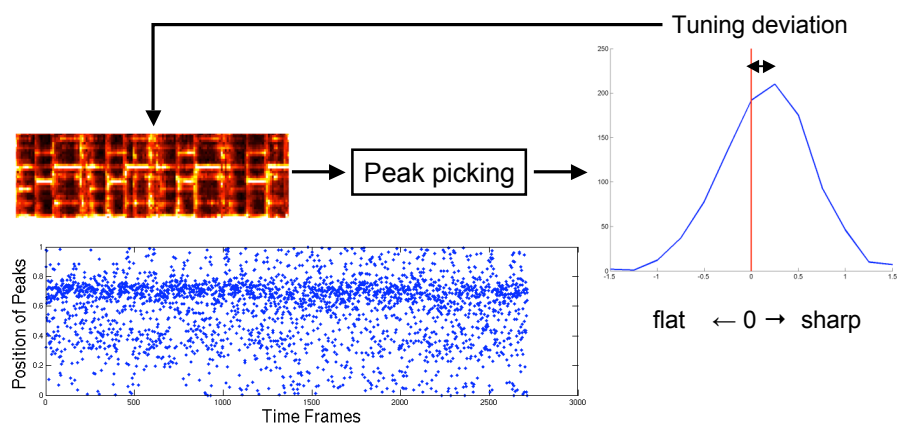
- We can create simple chord templates for common chords:
- E.g. Triads have a simple formulation of the form:
  - Major: **n, n + 4, n + 7** (e.g. G = [g, b, d])
  - Minor: **n, n + 3, n + 7** (e.g. g = [g, b^b, d])
- Use simple patterns with 1 on composing notes, 0 elsewhere
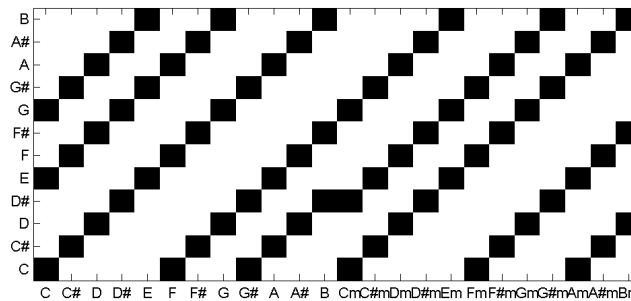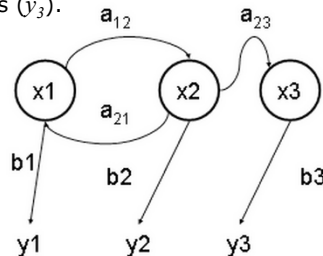

Chord Estimation (4)

- The maximum of the correlation between chord templates and 12-bin chromas represents the instantaneous chord value
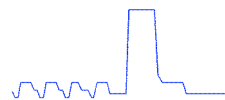
Chord Estimation (5)

- We can think of chords as the states of a process (x_3) and chromas as the result of that process (y_3).


- We can calculate the *probability* of observing certain chroma values for a given chord (b_3)
- Furthermore, states are not independent, but the occurrence of a certain state depends on the previous occurrence of other states
- The simpler of such probabilistic processes, a Markov process, is a random process where the probability of the occurrence of the current state (a_{23}) depends only on the occurrence of the previous one
- Moreover, my states are not directly observable: are hidden from me.

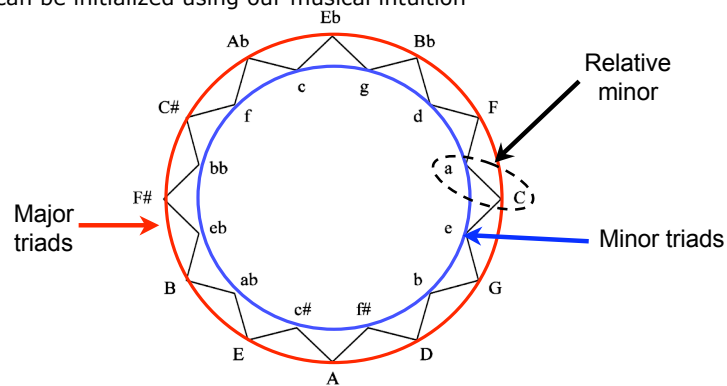# Chord Estimation (6)

Viterbi algorithm
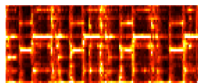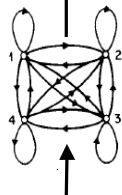
Tuned and Beat-synchronous chroma

- Ergodic HMM
- Chord space: e.g. major and minor triads
- Standard algorithms for the calculation of parameters
- Decoded using the Viterbi algorithm

Bello and Pickens, 2005

Chord Estimation (7)

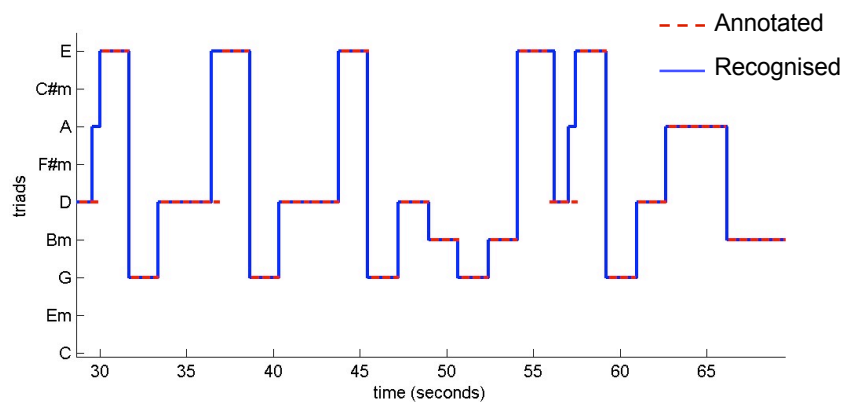- x_n are defined as a set of chords (e.g. 24 major and minor triads)
- y_n are the chroma vectors out of our analysis of the signal
- b_n can be initialized as simple chord templates (like before)
- a_{mn} can be initialized using our musical intuition

- All parameters can be efficiently estimated using the Baum-Welch method
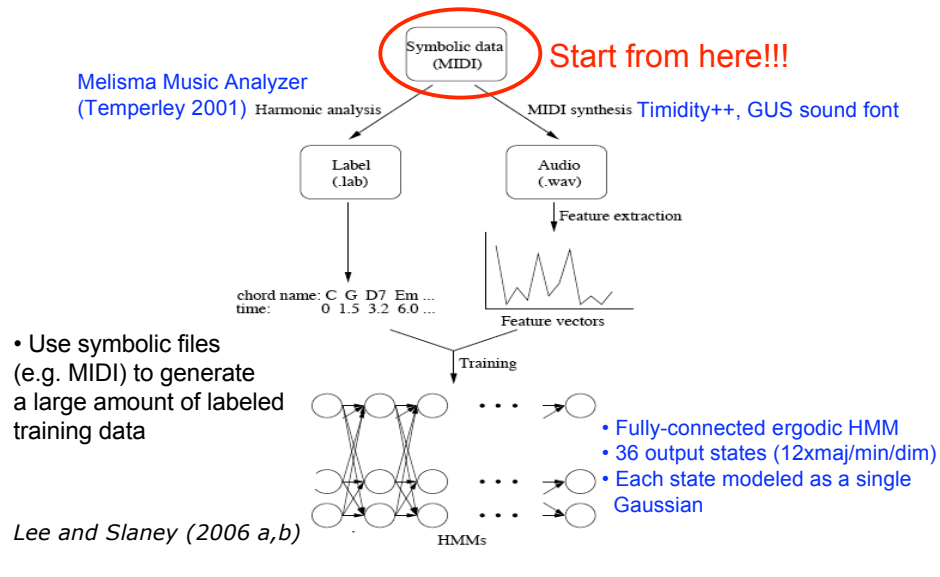

Chord Estimation (8)

- Eight days a week – The Beatles

Bello and Pickens (2005)

10

Chord Estimation (9)

Symbolic data (MIDI)   Start from here!!!

Melisma Music Analyzer
(Temperley 2001) Harmonic analysis   MIDI synthesis Timidity++, GUS sound font

Label (.lab)   Audio (.wav)

Feature extraction

chord name: C G D7 Em ...
time: 0 1.5 3.2 6.0 ...

Feature vectors

• Use symbolic files
(e.g. MIDI) to generate
a large amount of labeled
training data

Training

• Fully-connected ergodic HMM
• 36 output states (12xmaj/min/dim)
• Each state modeled as a single
  Gaussian

Lee and Slaney (2006 a,b)   HMMs

References

- Cabral, G. Briot, J.-P., Pachet, F. Impact of distance in pitch class profile computation. Proceedings of the 10th Brazilian Symposium on Computer Music (SBCM'2005). (2005).
- Brown, J.C. and Puckette, M.S.. An Efficient Algorithm for the Calculation of a Constant Q Transform, J. Acoust. Soc. Am. 92, 2698-2701. (1992)
- Carol Krumhansl. Cognitive Foundations of Musical Pitch. Oxford: Oxford University Press. (1990).
- Gómez, E. Herrera, P. Estimating The Tonality Of Polyphonic Audio Files: Cognitive Versus Machine Learning Modelling Strategies. Proceedings of the Fifth International Conference on Music Information Retrieval, Barcelona, Spain, 2004.
- Harte, C. Sandler, M. Automatic chord identification using a quantised chromagram. Proceedings of 118th AES Convention. Barcelona. (2005).
- Bello, J.P. Pickens, J. A robust mid-level representation for harmonic content in music signals. Proceedings of the 6th International Conference on Music Information Retrieval. London. (2005).
- Lee, K, Slaney, M, ". Proceedings of the 7th International Conference on Music Information Retrieval, Victoria (2006).
- Lee, K, Slaney, M, ". Published at the Audio and Musical Computing for Multimedia Workshop, Santa Barbara, California. (2006).