Multimedia Systems

IV. Music Information Retrieval

4.2. Pitch

António Sá Pinto
FEUP – 2019/2020

Agenda

- Pitch Detection
- Melody Estimation
- Polyphonic Transcription
- Chroma, Chords and Keys
- Applications

Overview

Pitch

f0 estimation

What frequency is that sound?

harmonic mixing

How can I mix these songs together?

melody estimation

Which notes am I humming?

polyphonic transcription

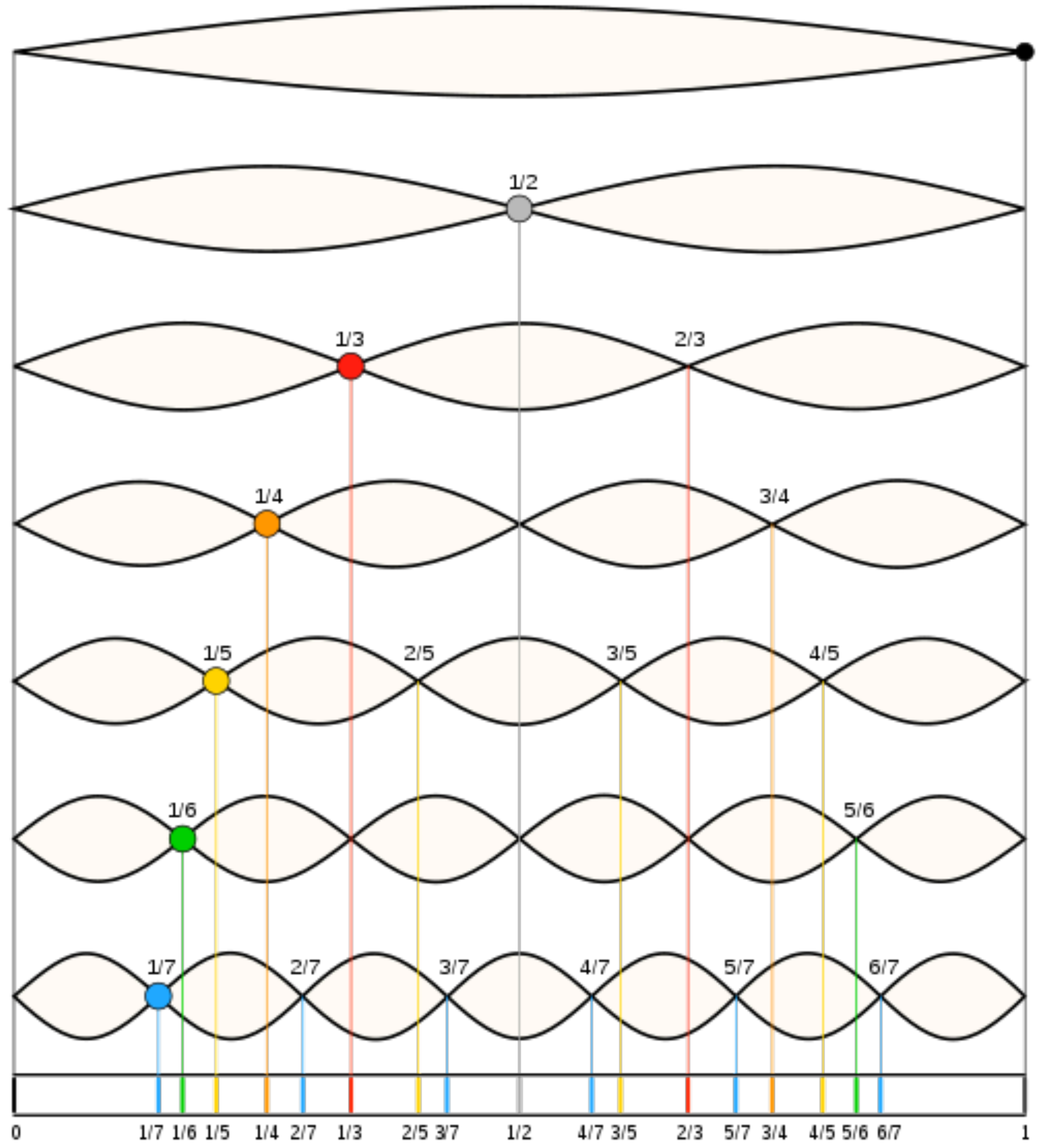
Which notes are playing?

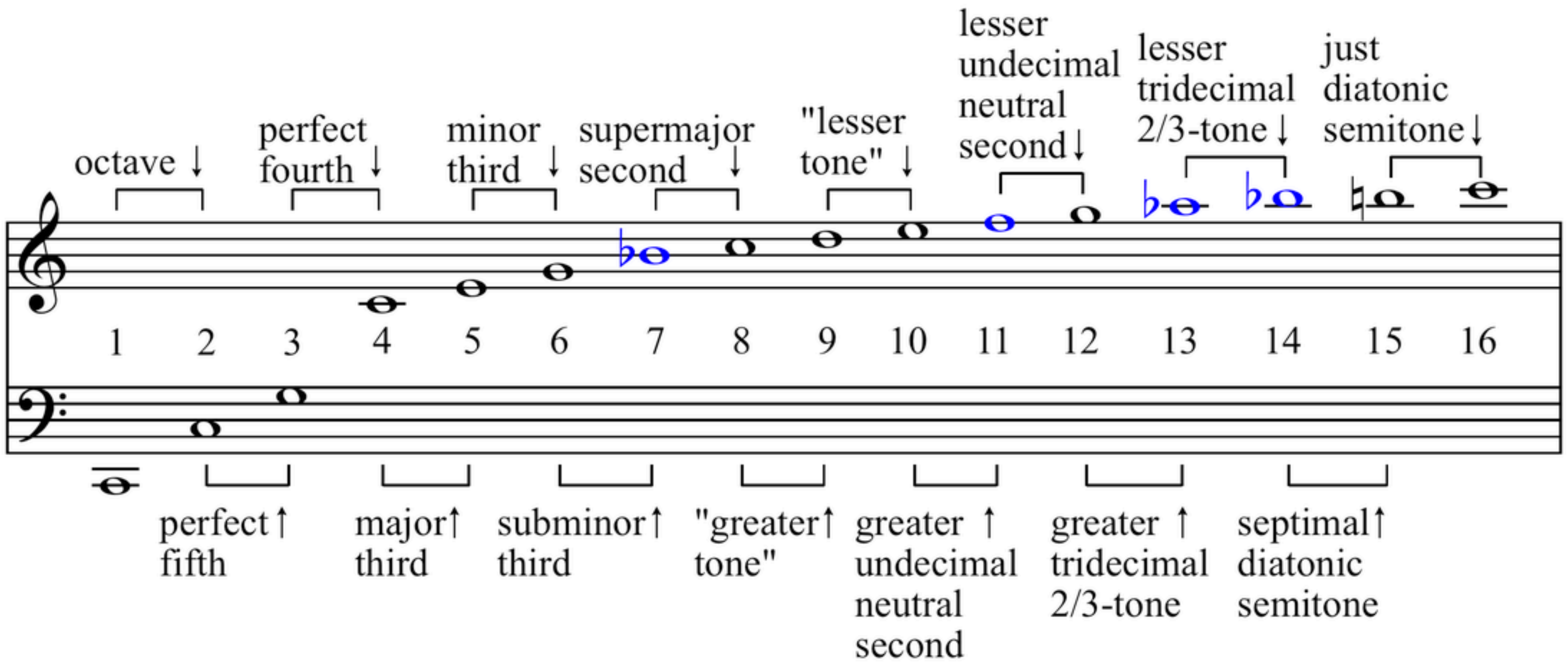
How can I play the chords in this song?

chord detection

Sound

Pure tones, fundamental, partials, harmonics, overtones, inharmonics.

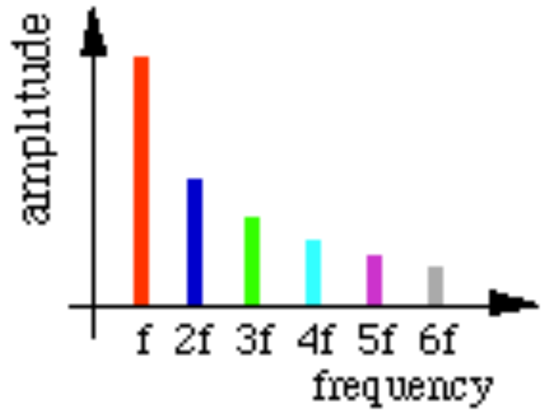


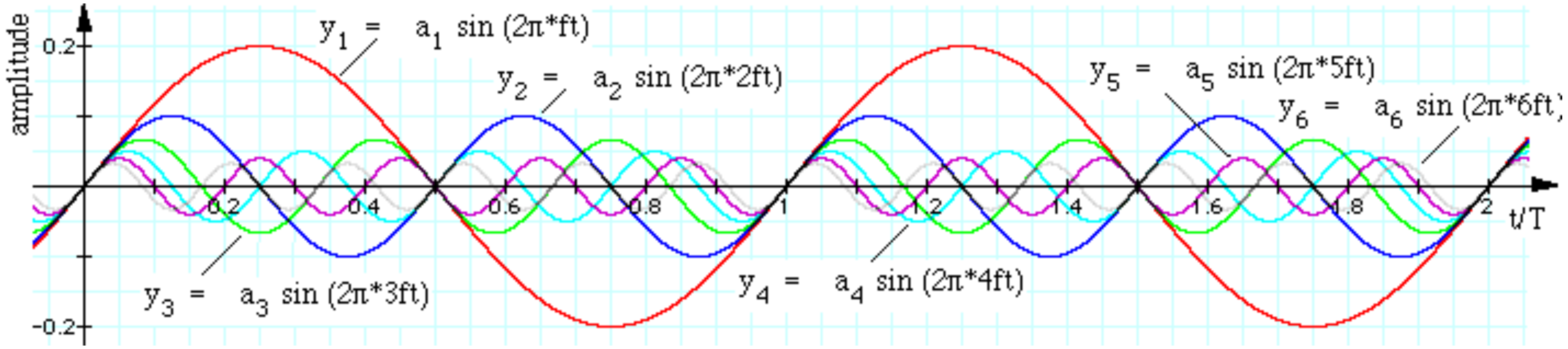


https://en.wikipedia.org/wiki/Harmonic_series_(music)

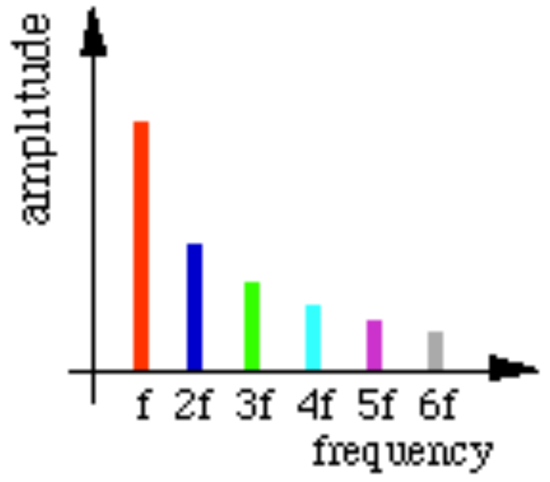
Sound

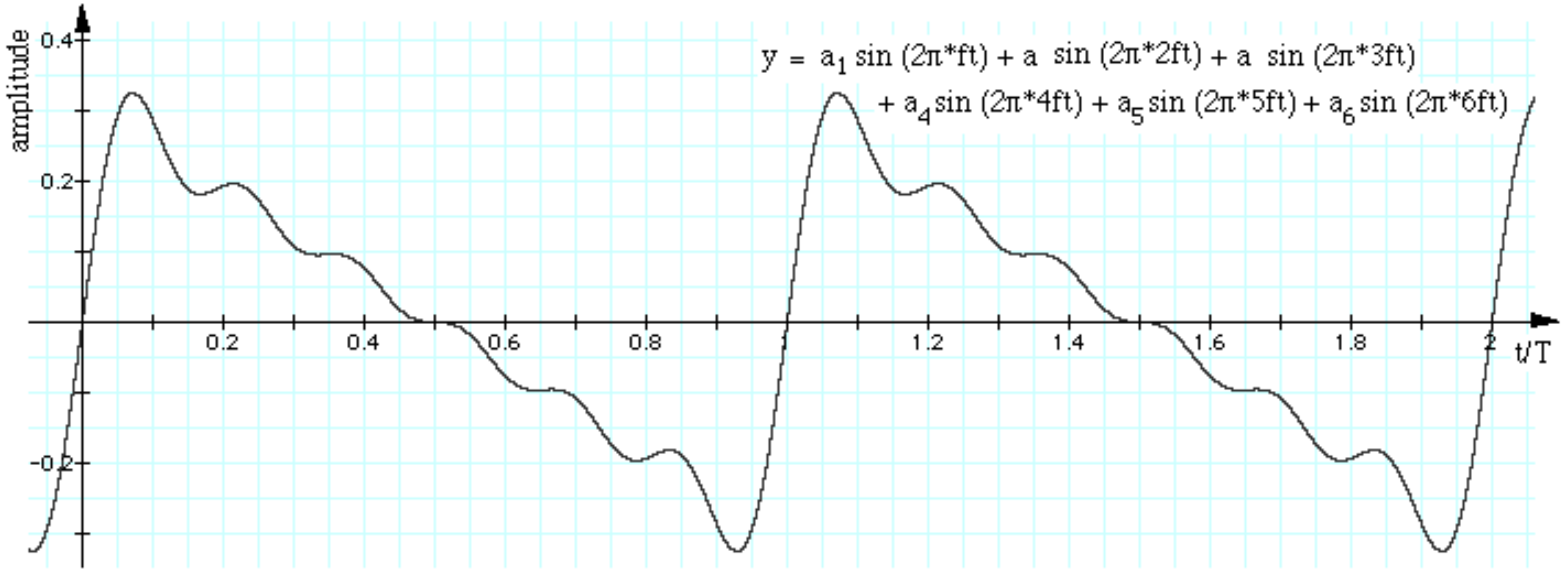
Pure tones, fundamental, partials, harmonics, overtones, inharmonics.





The harmonic series is an arithmetic series (1×f, 2×f, 3×f, 4×f, 5×f, …).
The components of a (complex) tone which are not multiples of f0 are called inharmonics.

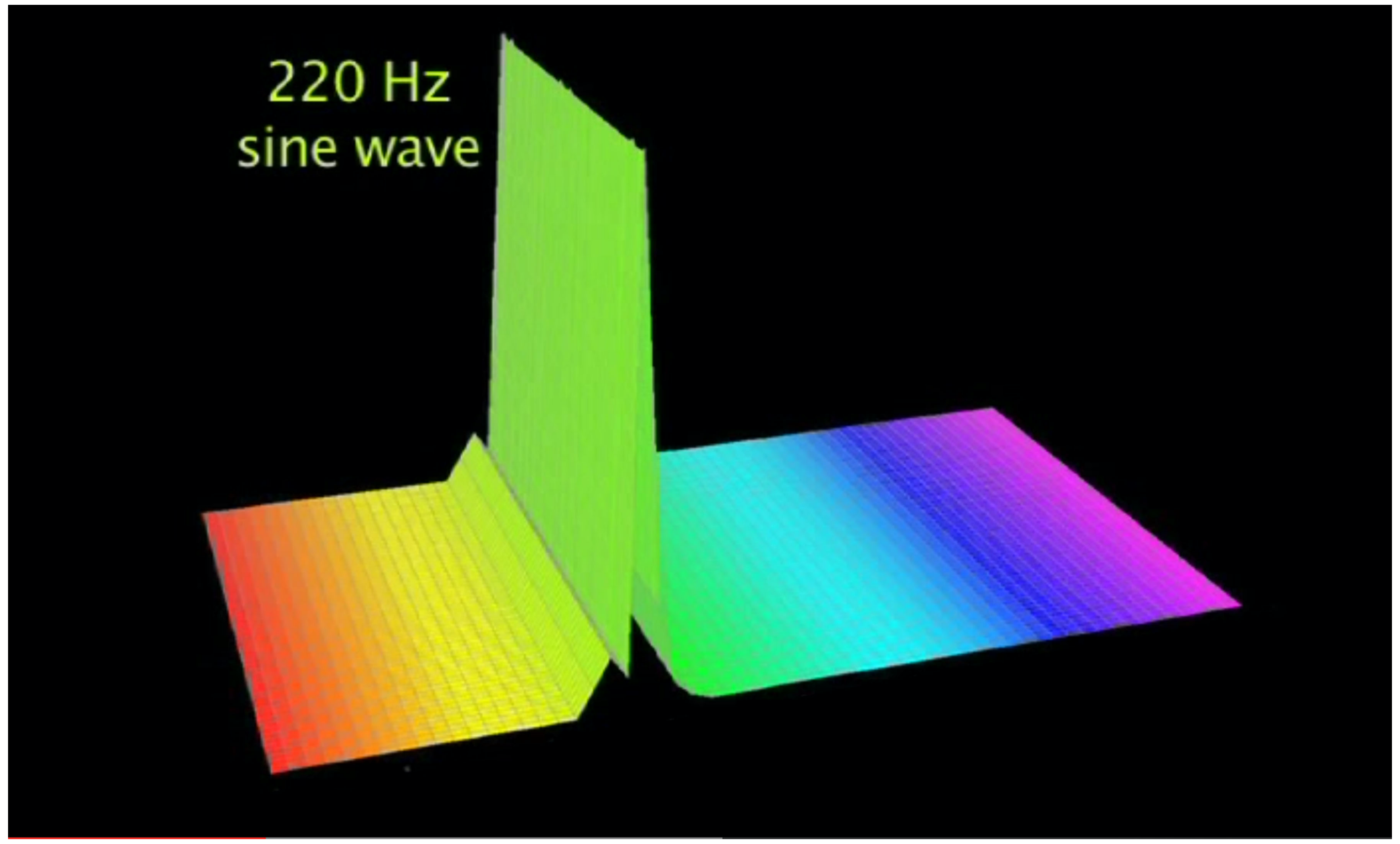


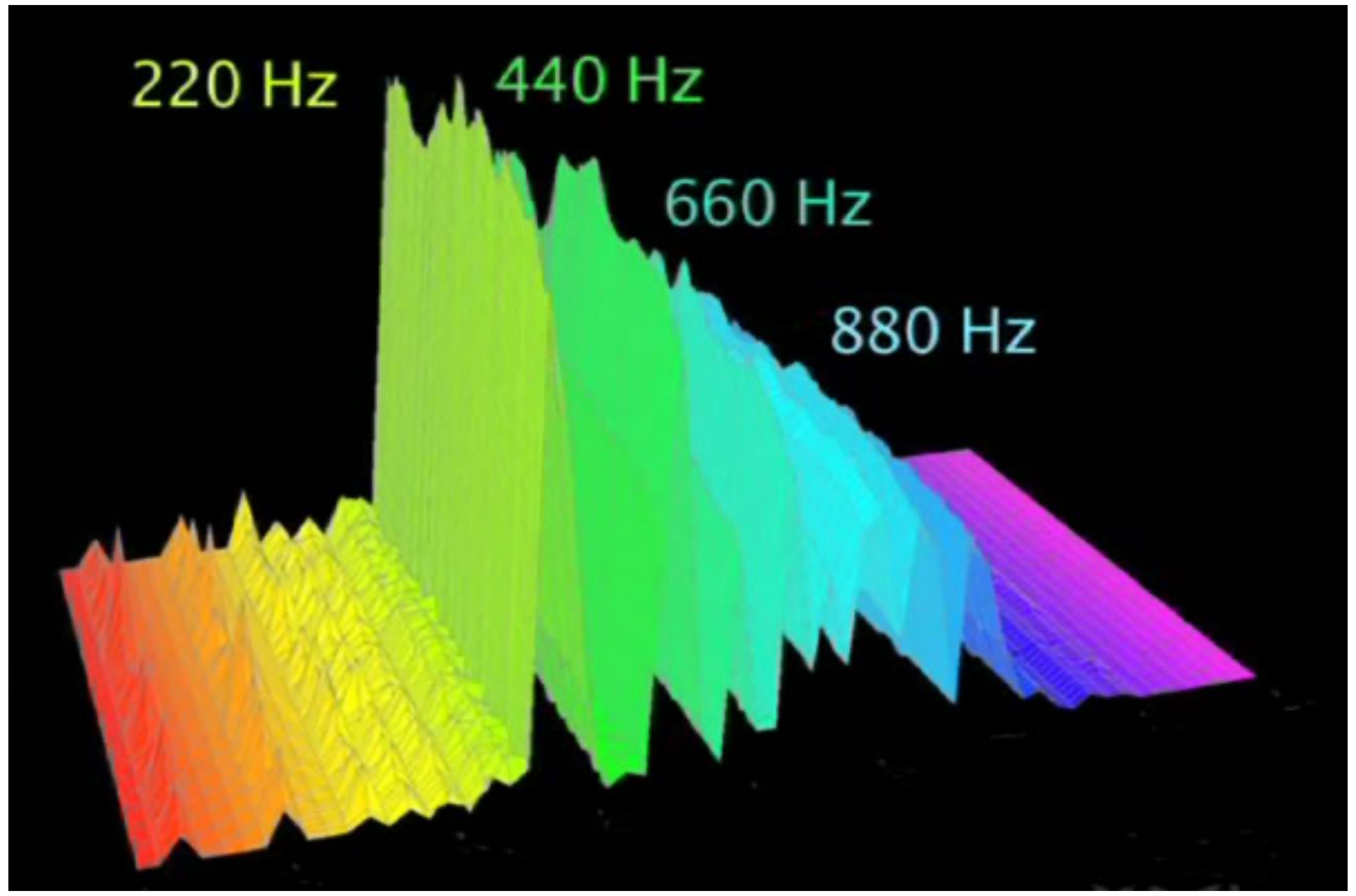


From: https://newt.phys.unsw.edu.au/jw/musical-sounds-musical-instruments.html

Sound

Pure tones, fundamental, partials, harmonics, overtones, inharmonics.





[From: https://learn.dojouniversity.com/understanding-harmonics-part-1/](https://learn.dojouniversity.com/understanding-harmonics-part-1/)

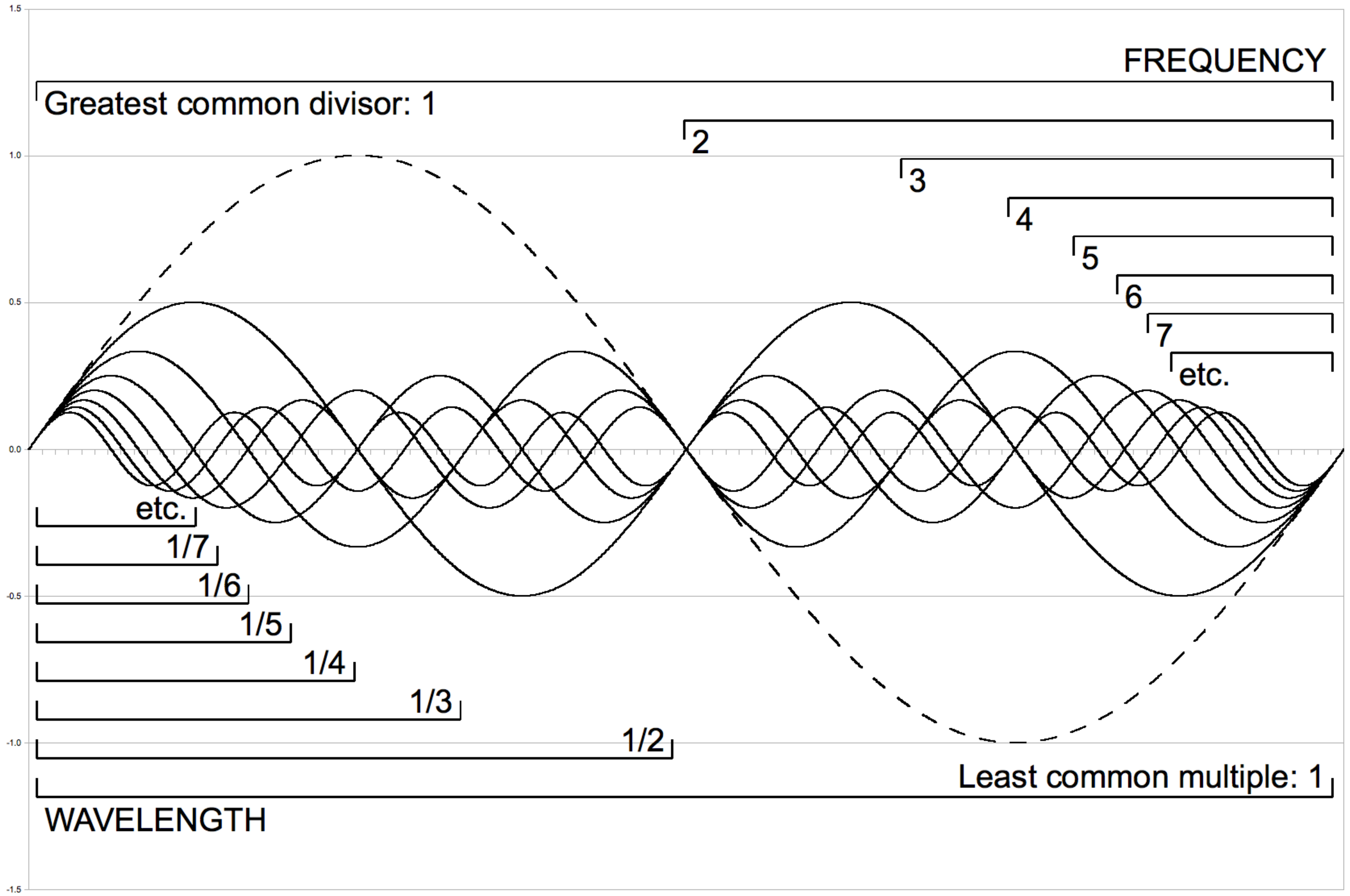
What is Pitch?

- Pitch is the quality that makes it possible to judge sounds as "higher" and "lower" in the sense associated with musical melodies
- Pitch may be quantified as a frequency, but pitch is not a purely objective physical property; it is a subjective psychoacoustical attribute of sound
- In general, pitch is given by the lowest frequency of a sound, known as the **fundamental frequency**, f_0

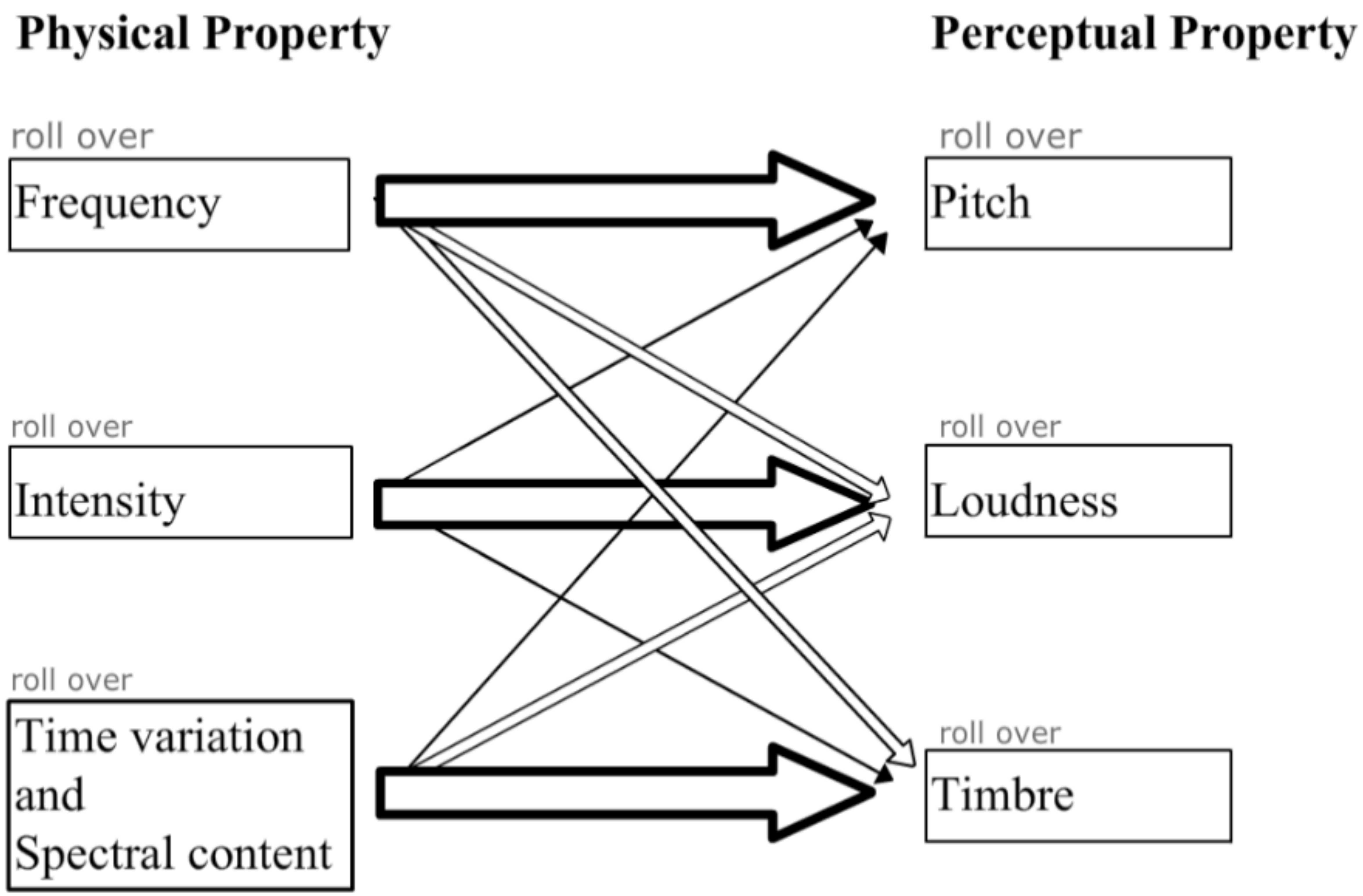
https://en.wikipedia.org/wiki/Pitch_(music)

Pitch

Pitch is (roughly) given by the greatest common divisor frequency of a sound, known as the **fundamental frequency**, f_0



Pitch



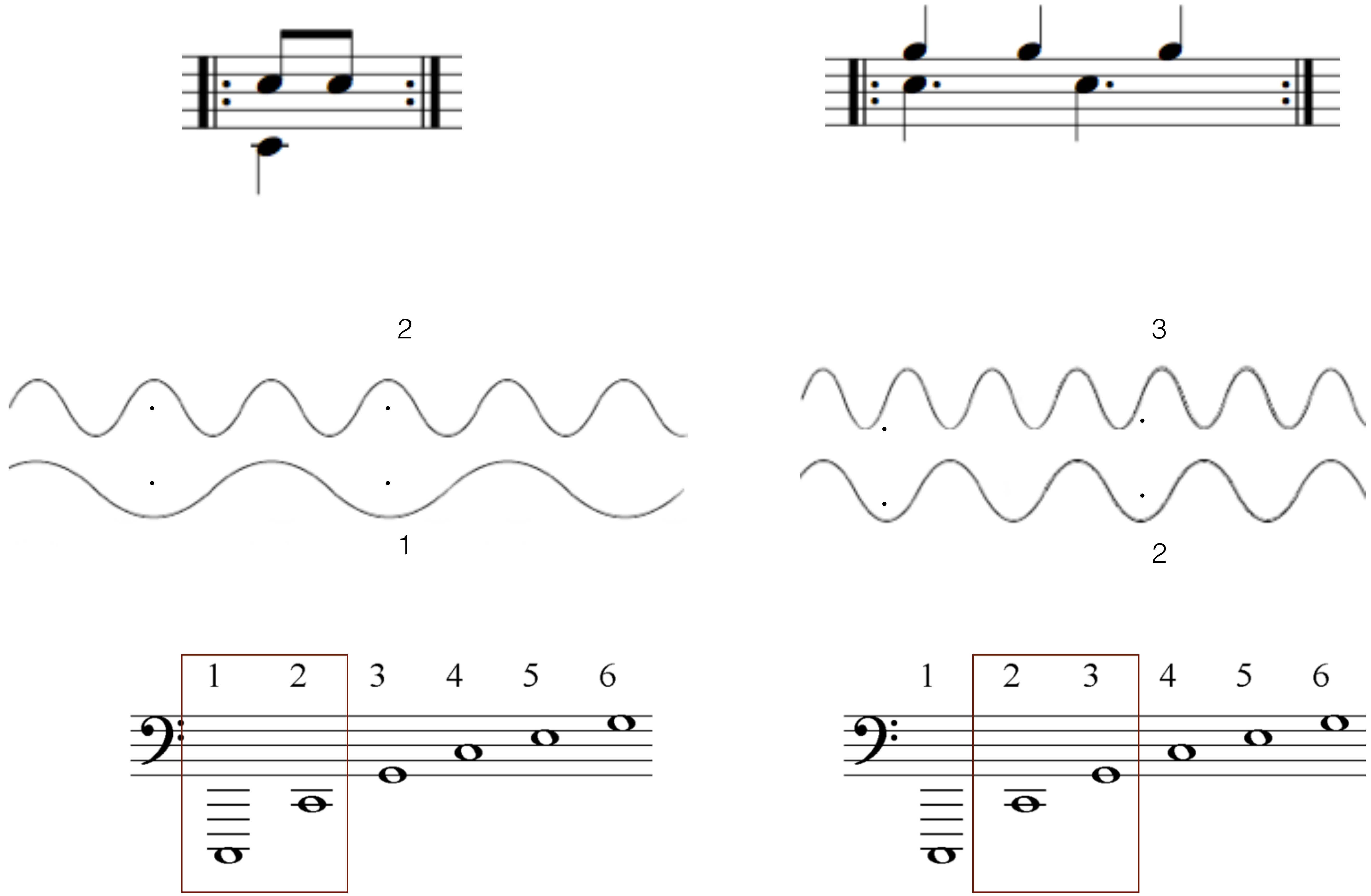
from: (UNES.EDU) Introduction to Sound

Pitch/Rhythm Duality

- periodicities > 20Hz are perceived as **pitched**
- periodicities < 20Hz are perceived as **rhythmic**

Let’s hear 2 examples

Pitch/Rhythm Duality



From: http://dantepfer.com/blog/?p=277

Pitch and Rhythm

Rhythm - Pitch duality

Time - Frequency duality

Fourier Transform

Auditory Illusion

Shepard Tone is a sound consisting of a superposition of sine waves separated by octaves. When played with the bass frequency of the tone moving upward or downward, it creates the auditory illusion of a tone that continually ascends or descends in pitch, yet which ultimately seems to get no higher or lower. This continuous version is called the Shepard-Risset Glissando.

https://en.wikipedia.org/wiki/Shepard_tone

Task2: *ExplainMe* sound

GOAL

- Bi-dimensionality and circularity of pitch perception.

(Escher) Relativity 1953

Auditory Illusion

A very known use of Shepard Tone.

Auditory Illusion

FMP Notebooks

- https://www.audiolabs-erlangen.de/resources/MIR/FMP/C1/C1S1_ChromaShepard.html

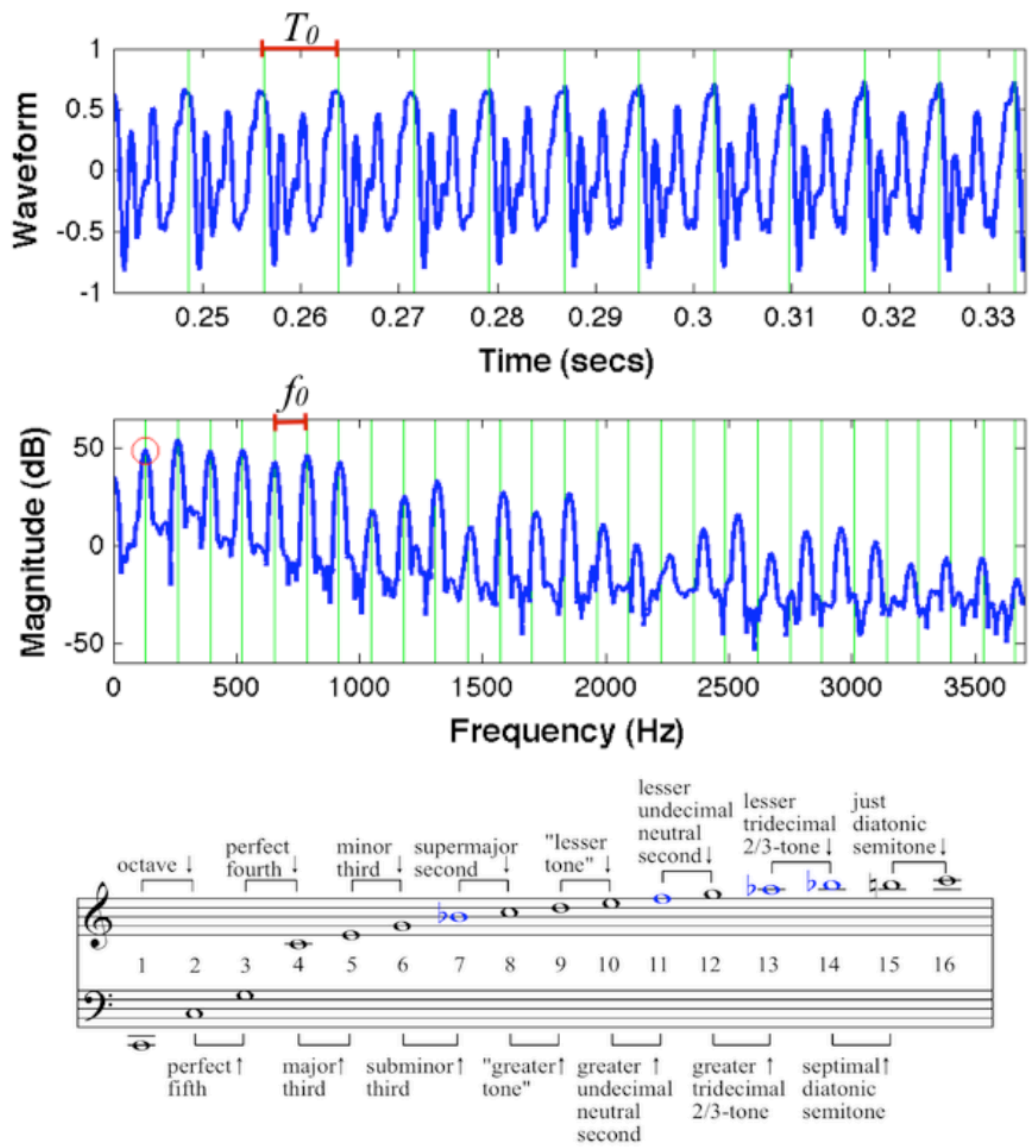
- Generate Shepard Tones and Shepard-Risset Glissando!

Pitch Estimation

- Monophonic vs
- Polyphonic music signals: *two or more notes sounding simultaneously, be it different instruments (e.g. voice, guitar and bass) or a single instrument capable of playing more than one note at a time (e.g. the piano)* (Salamon et al. 2013)

Monophonic Pitch Estimation (f_0)

Periodic (time-domain) - harmonic (frequency domain: $f_0, 2 \cdot f_0, 3 \cdot f_0, \ldots$)



Monophonic Pitch Estimation

A large number of approaches (>100)

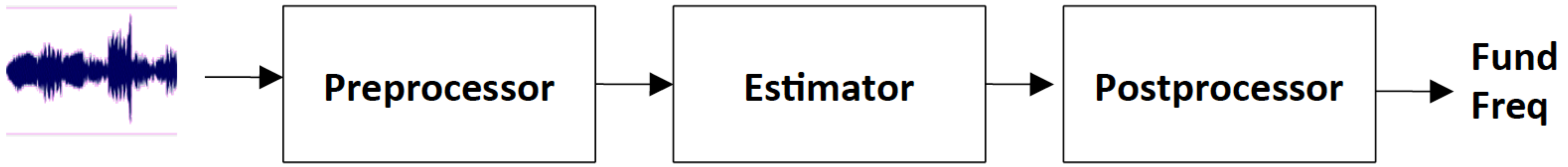
- Time vs Frequency domain
- Adapted to different instruments

Difficulties

- Quase-periodicities
- Multiple periodicities associated with a given f_0
- Noise
- Etc.

Monophonic Pitch Estimation

- Three main steps in f0 estimation



- Preprocessor: data reduction, noise reduction.
- Estimator/Extractor:
  - Time-domain algorithms
  - Frequency-domain algorithms
- Postprocessor: error correction, smoothing, tracking, etc.

Monophonic Pitch Estimation

Methods to detect periodicities

??

Monophonic Pitch Estimation

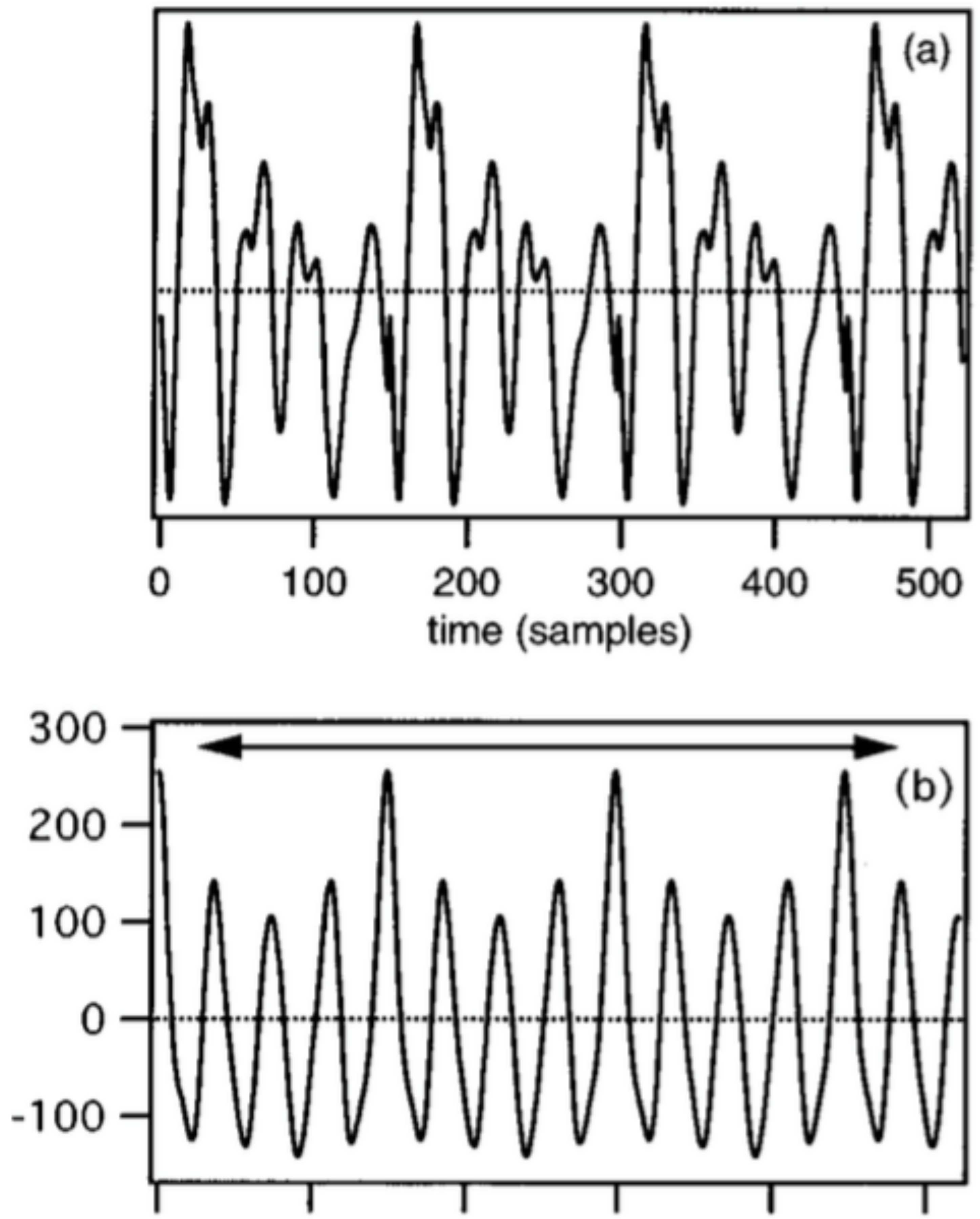
Methods to detect periodicities

Zero-crossing rate

DFT

Autocorrelation (Rabiner, 1977)

Etc.

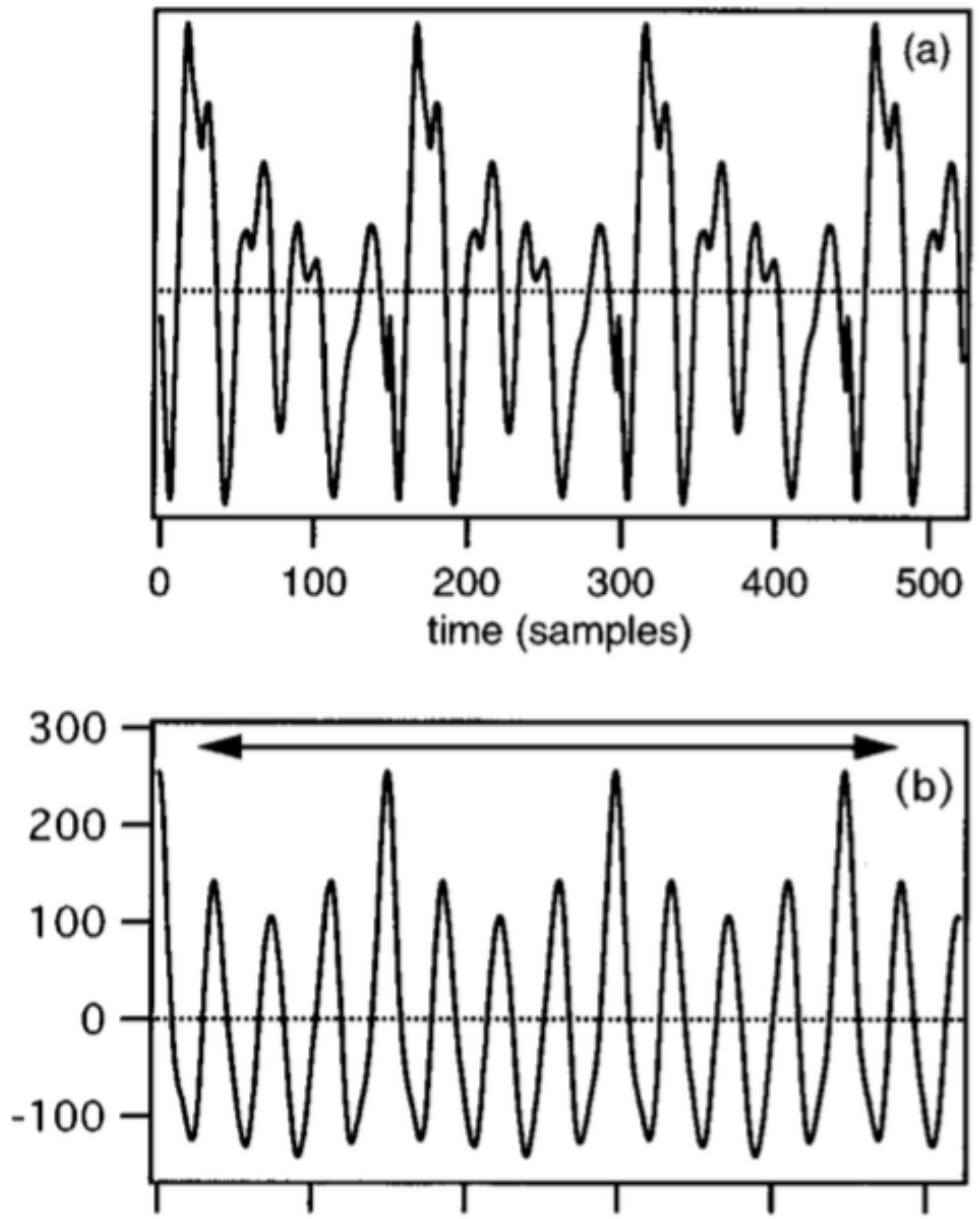


Monophonic Pitch Estimation

Detect periodicity

Zero-crossing rate

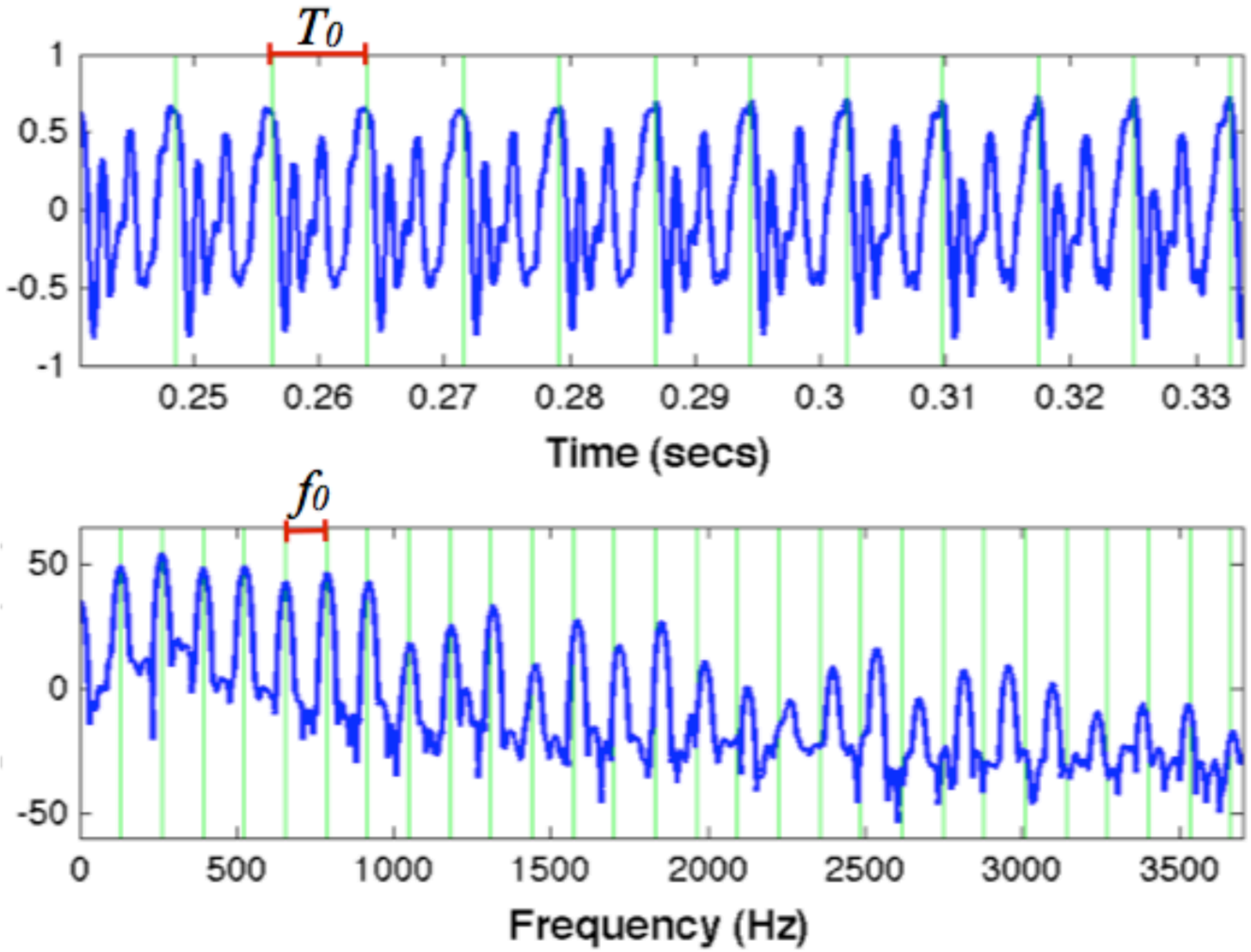
- Very simple & computationally inexpensive.
- Not very accurate



Monophonic Pitch Estimation

Detect periodicity

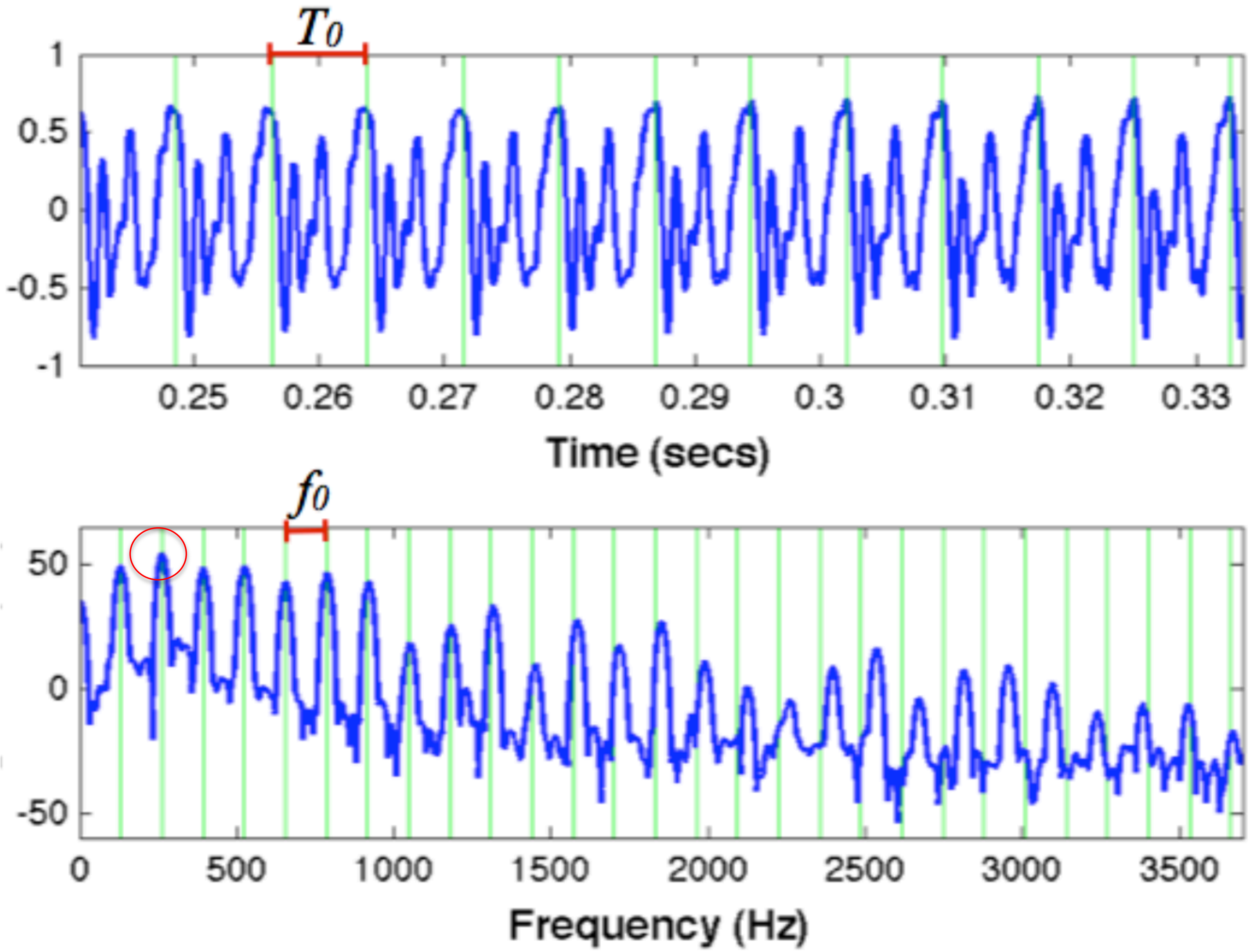
DFT



Monophonic Pitch Estimation

Detect periodicity

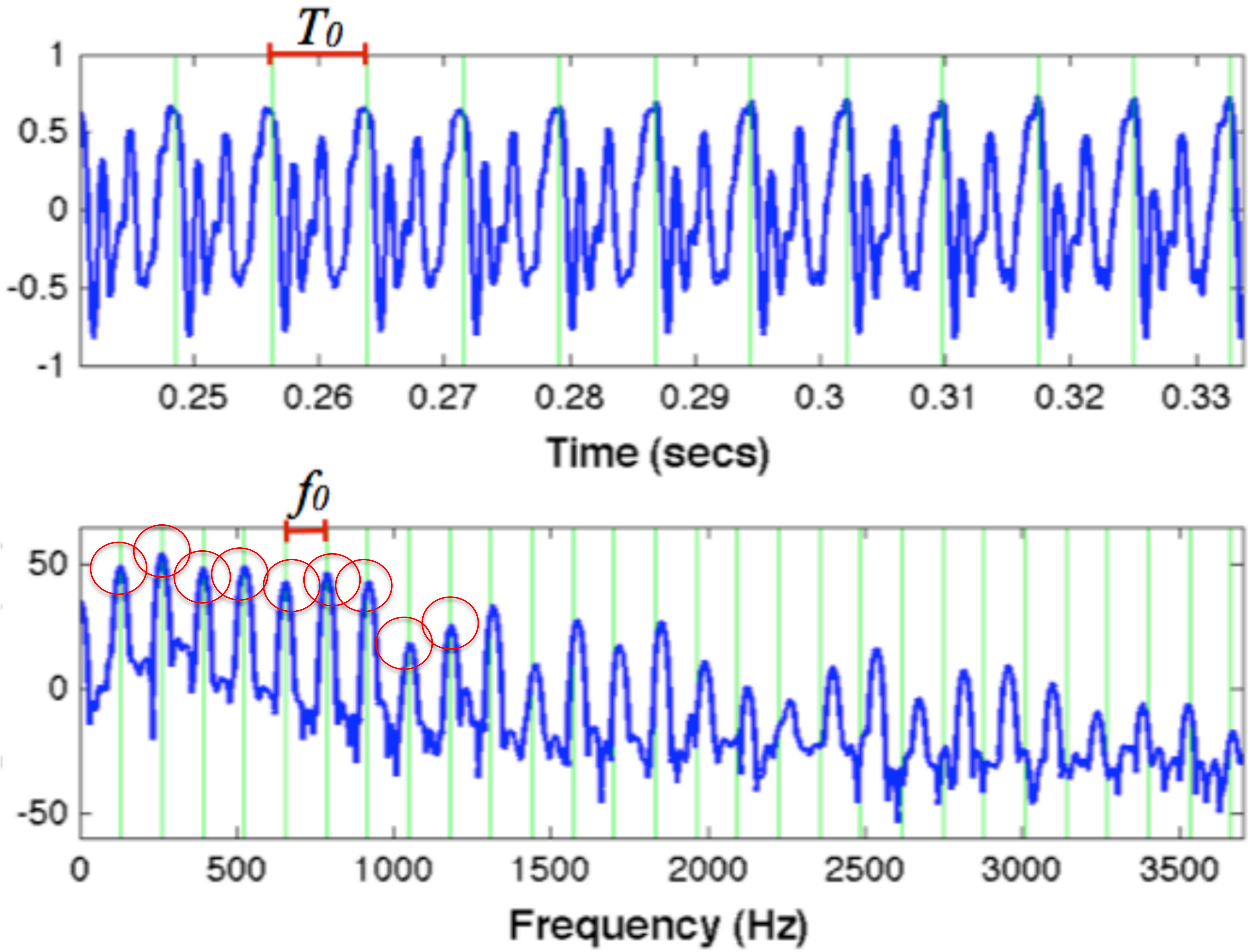
DFT



Monophonic Pitch Estimation

Detect periodicity

DFT

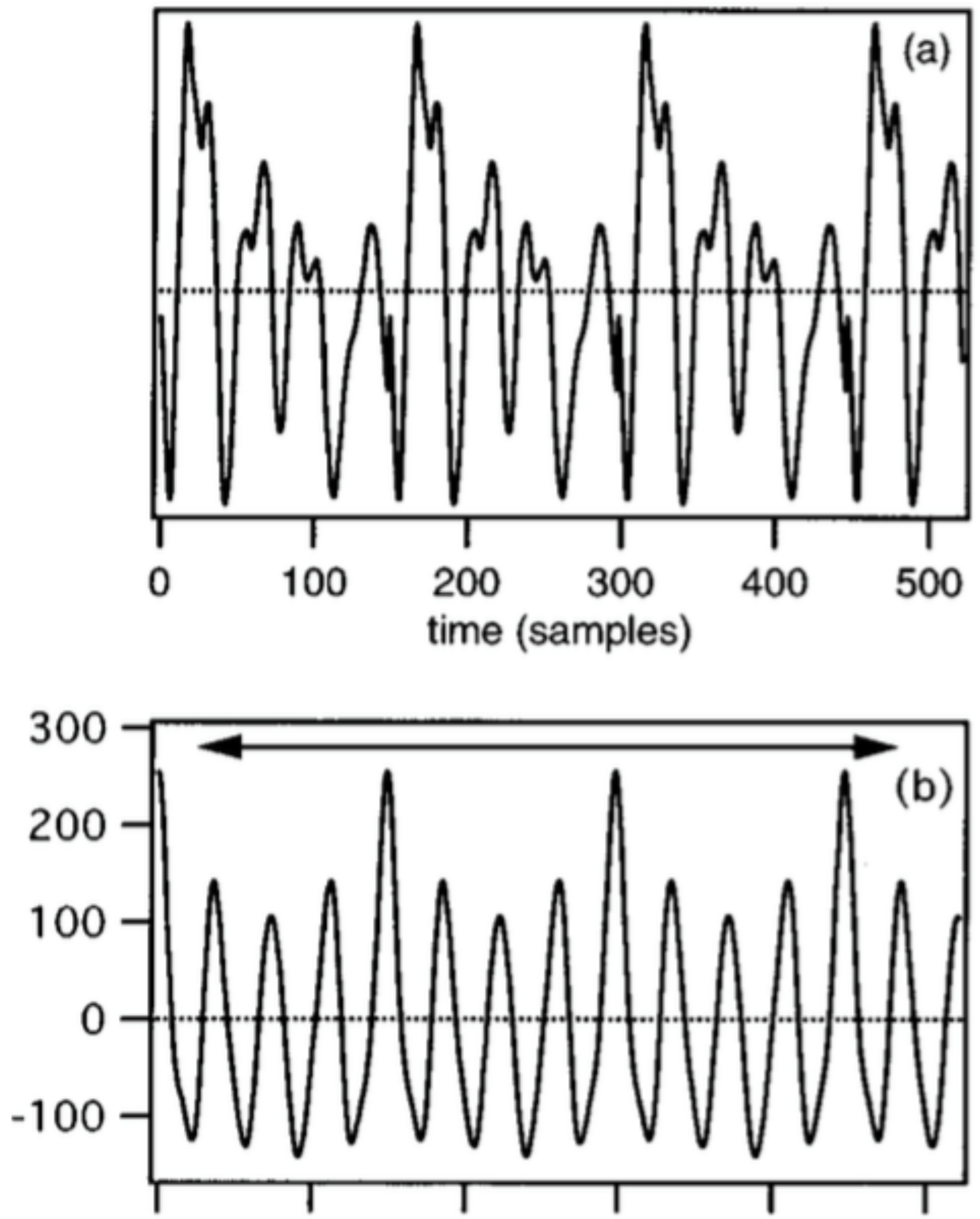


Monophonic Pitch Estimation

Detect periodicity

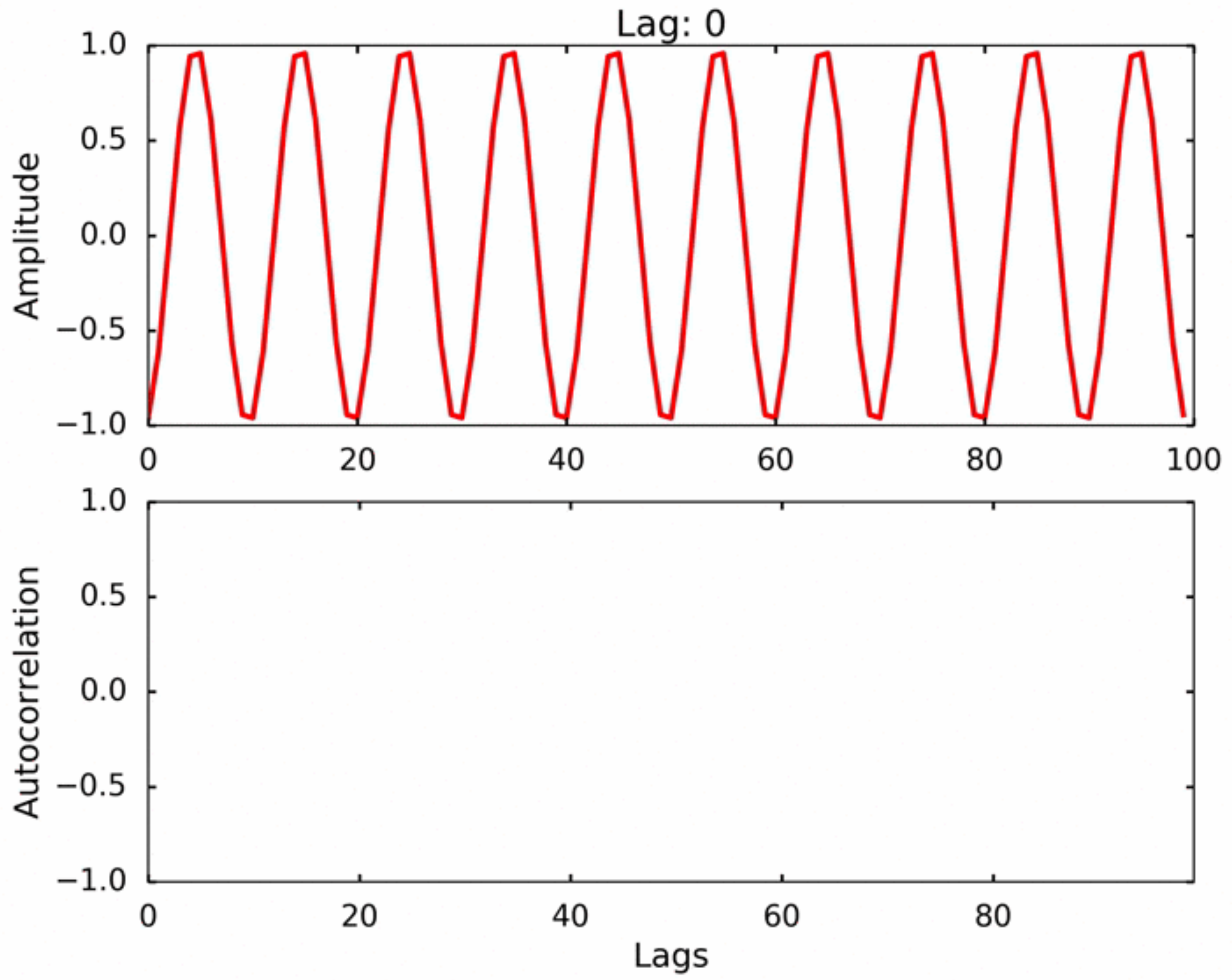
Autocorrelation (Rabiner, 1977)

- Cross-correlation is a measure of similarity of 2 signals across different lags (delays)
- Compare the signal against itself (auto) over a range of lags.
- Very intuitive!



Autocorrelation function (ACF)

Compare the signal to a shifted version of itself… for all possible shifts (time lags)



Monophonic Pitch Estimation

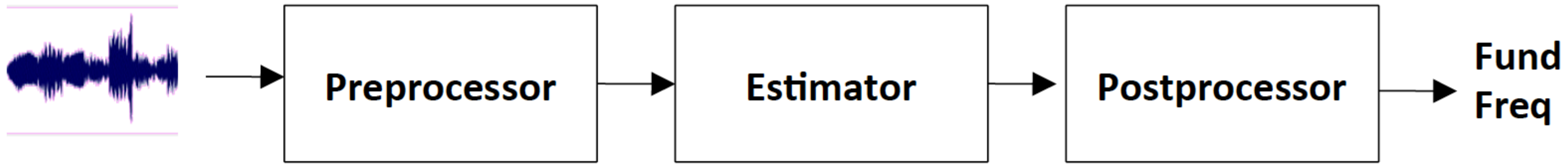
Other methods

AMDF

Cepstrum

YIN

etc.



Monophonic Pitch Estimation

Other methods

- AMDF
- Cepstrum
- YIN
- etc.

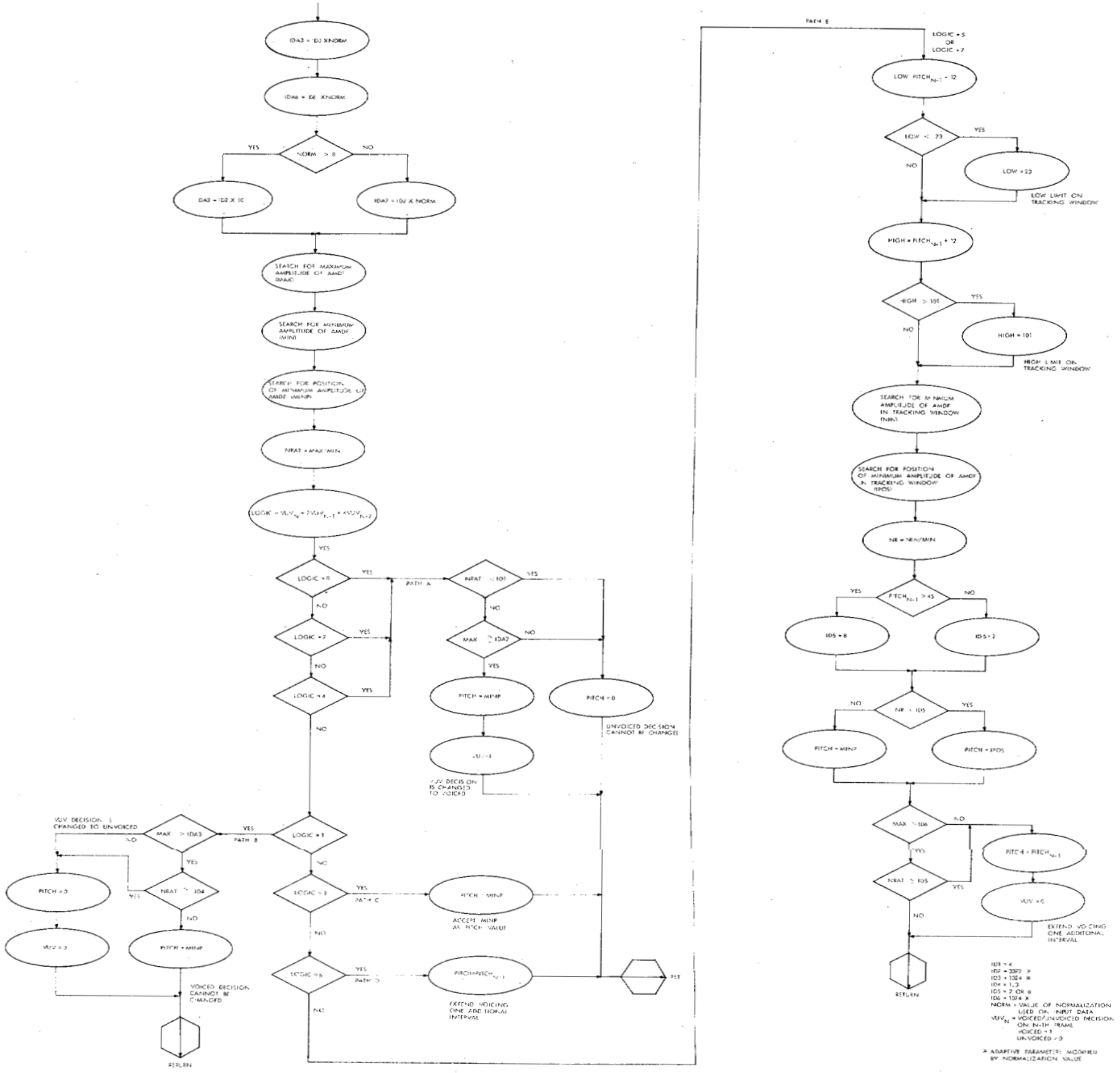


Fig. 7. AMDF pitch extraction logic flow chart.

Query By Humming

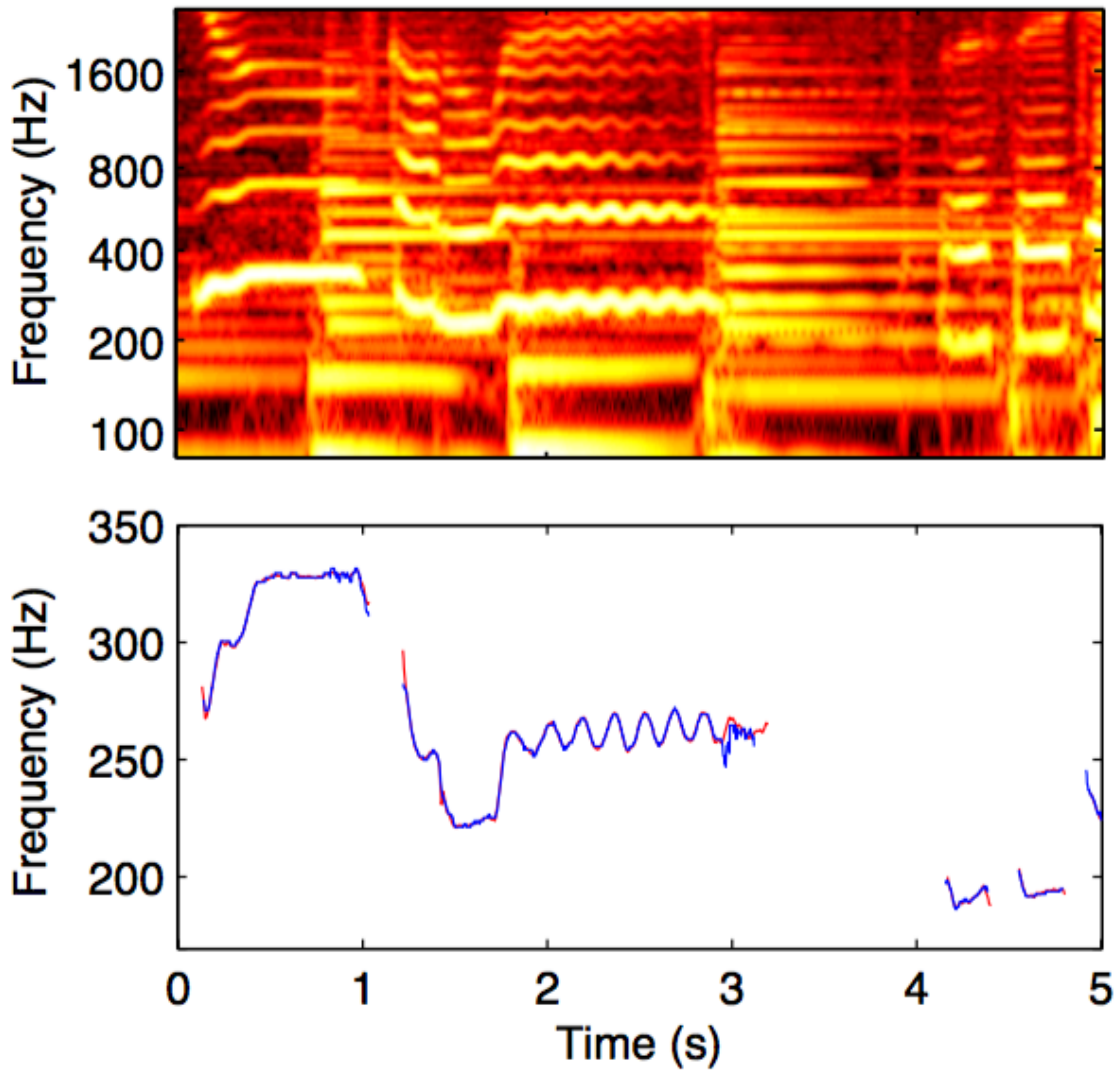
Tell me which song I'm humming

- **One of the classic problems** in music information retrieval, dating back to the beginning of the ISMIR conference in 1999
  - Run a monophonic pitch tracker
  - Quantise pitches to notes (create MIDI file)
  - Compare against a database of known melodies (N-gram approach)
  - Create a ranked list of matches
  - Online services available: http://www.midomi.com

Melody Estimation

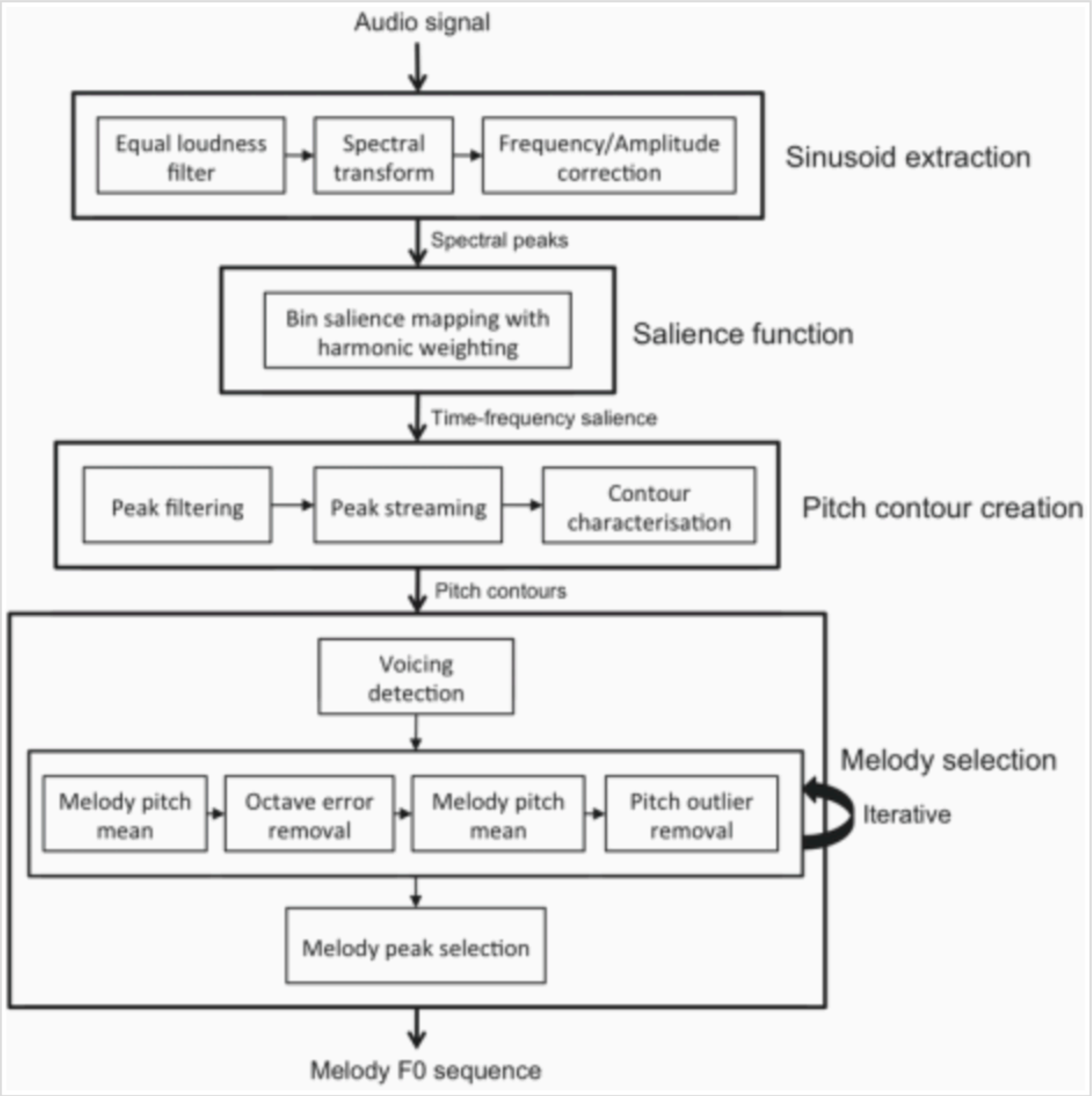
- Highly related to monophonic pitch estimation, but attempt to extract the melody **within a musical mixture**
- Creation of a melodic **salience** function - use reinforcement of harmonic spectral peaks
- Voiced/unvoiced detection - which notes are melody and which aren't

http://www.justinsalamon.com/melody-extraction.html#demo



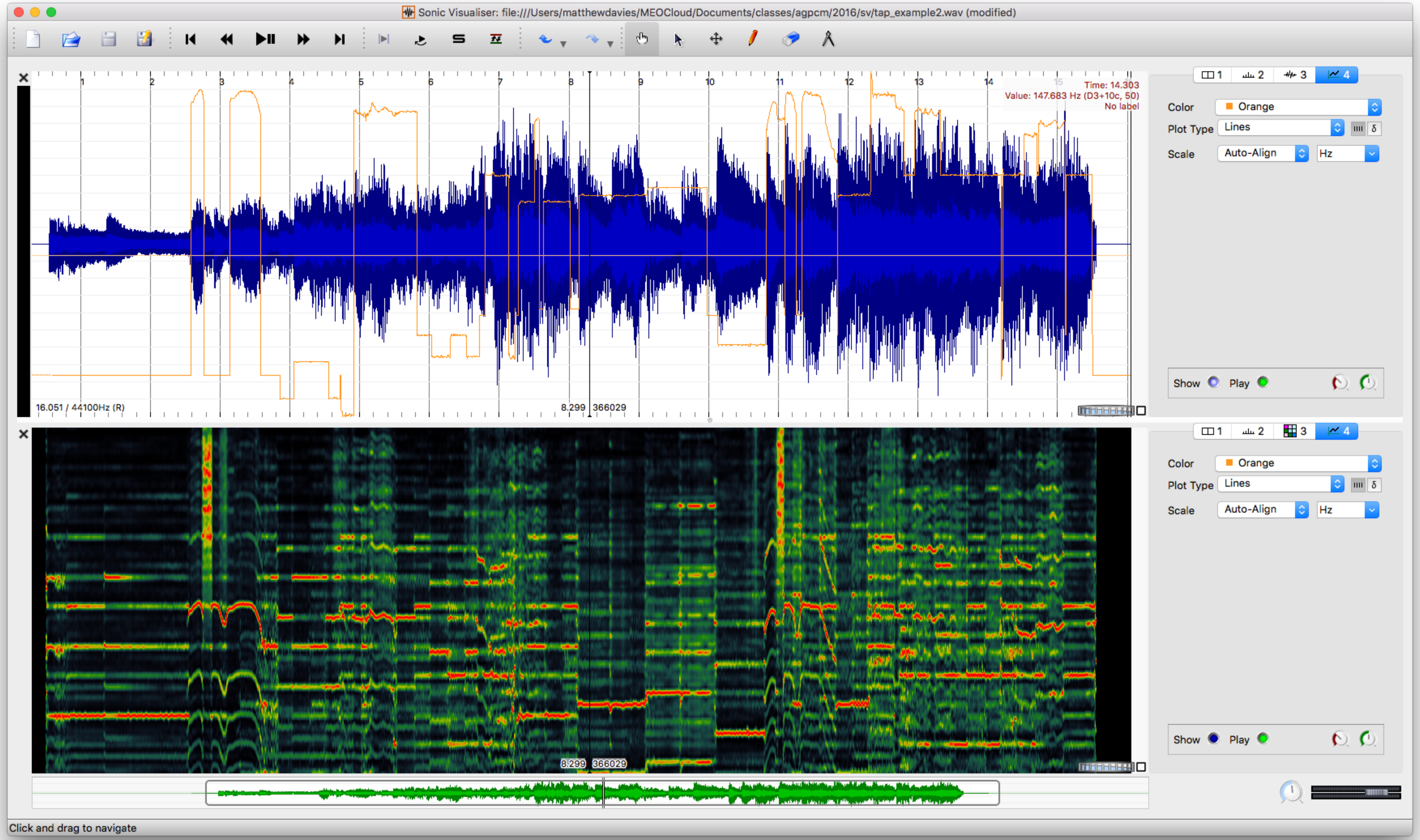
(Salomon, Gómez, Ellis, Richard), "Melody Extraction from Polyphonic Music Signals: Approaches, Applications and Challenges", IEEE Signal Processing Magazine, 31(2):118-134, Mar. 2014.

Melody Estimation



(Salomon, Gómez, Ellis, Richard), "Melody Extraction from Polyphonic Music Signals: Approaches, Applications and Challenges", IEEE Signal Processing Magazine, 31(2):118-134, Mar. 2014.

Sonic Visualiser: Melodia Plugin (VAMP)



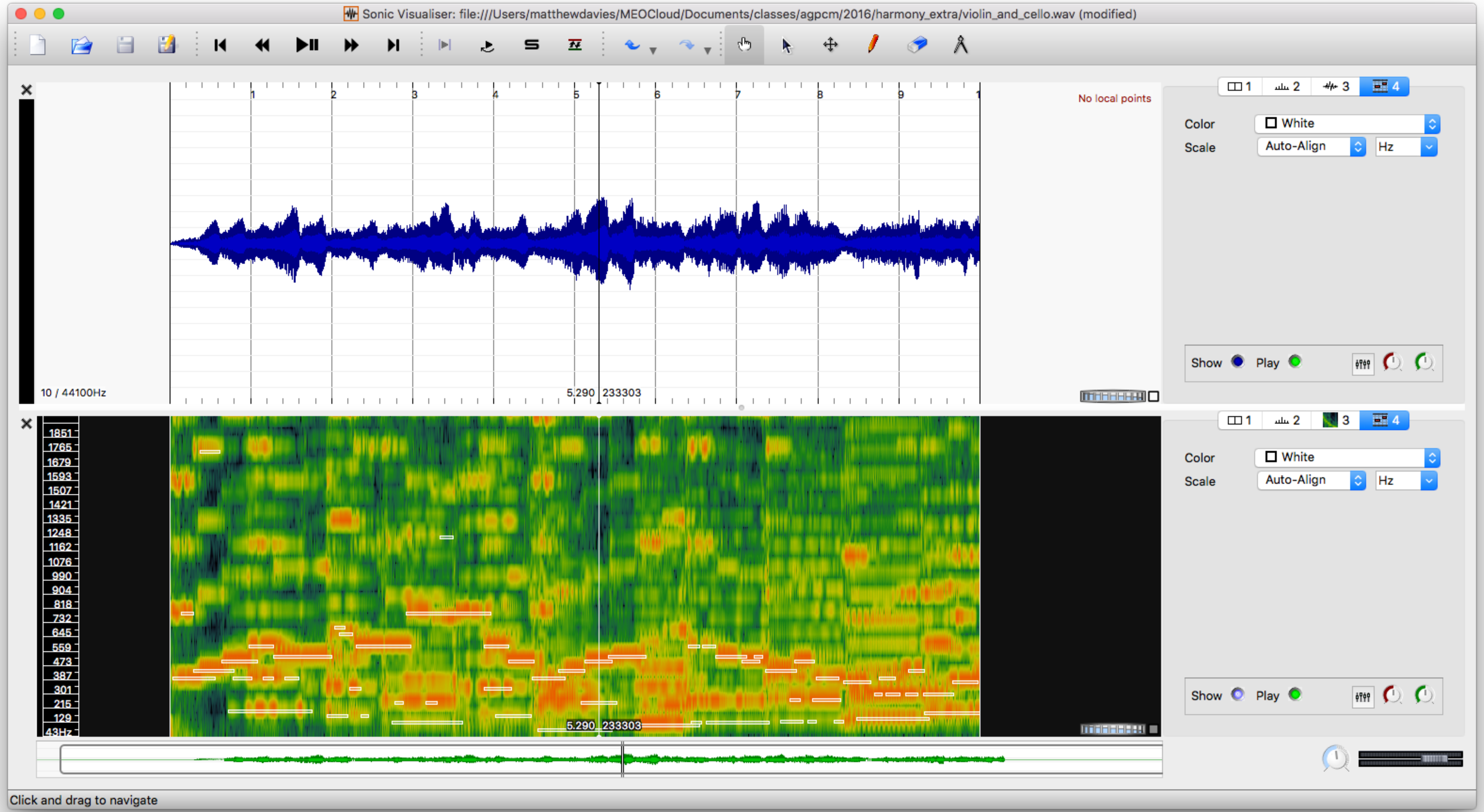
http://mtg.upf.edu/technologies/melodia

Polyphonic Transcription

If we can detect one note at a time, can't we detect multiple notes at once?

- Much harder problem! For humans too…
- Still considered one of the grand challenges of MIR
- Many issues in relation to overlapping frequency information from different sources
  - Detection of note onsets and offsets
  - Interference from percussive content
- Main assumptions: harmonicity, spectral smoothness, constant spectral template

Polyphonic Transcription (Silvet VAMP Plugin)



https://code.soundsoftware.ac.uk/projects/silvet/files

www.cs.tut.fi/~klap/iiro/

pitch summary

- Monophonic pitch tracking is among the easier problems (can be done in real-time - think *autotune*)
- Query by humming is mature enough to be commercially viable (*but maybe not for our "not-so-tuned" voices*)
- Melody detection is hard - very difficult to identify what is singing voice compared to other instruments with vibrato
- Polyphonic Transcription is very very hard;

4.2. Pitch

.

Pitch (Chroma and Height)

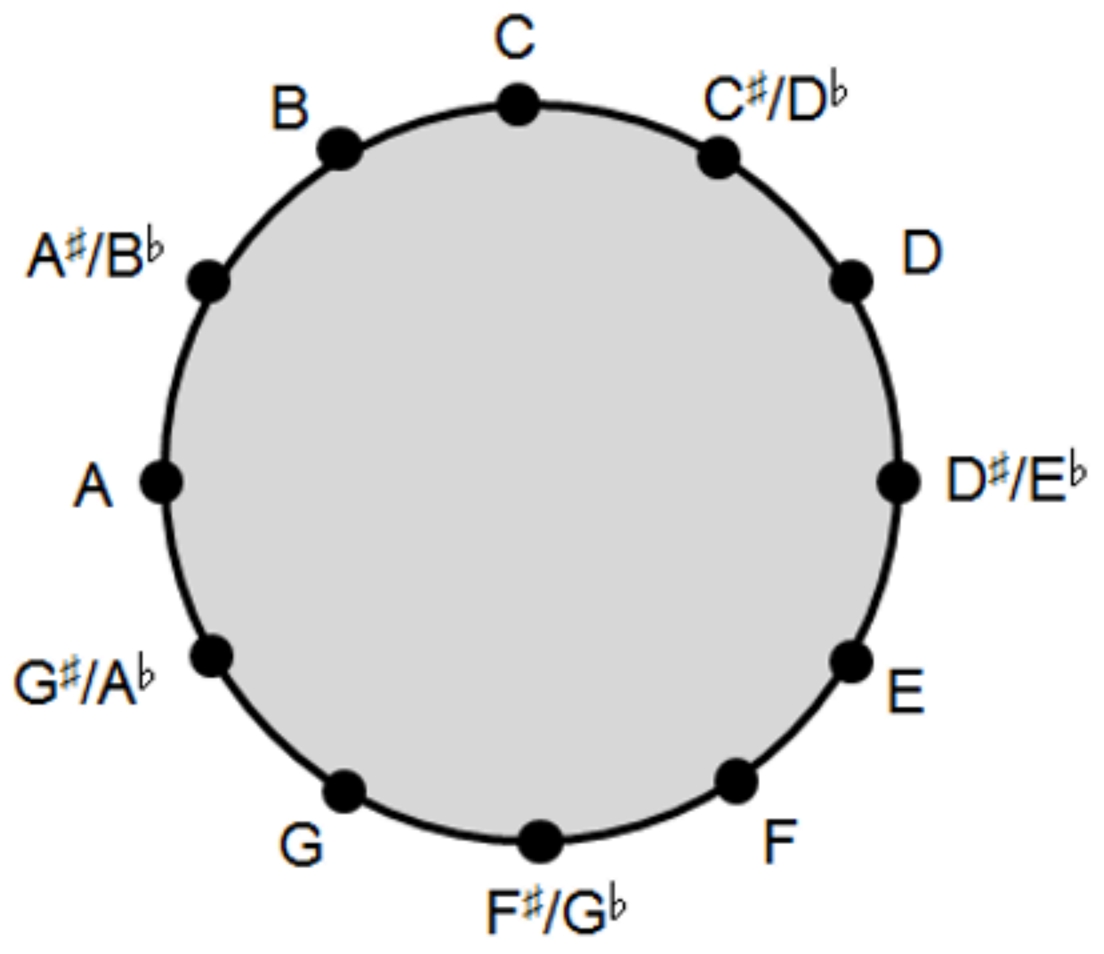
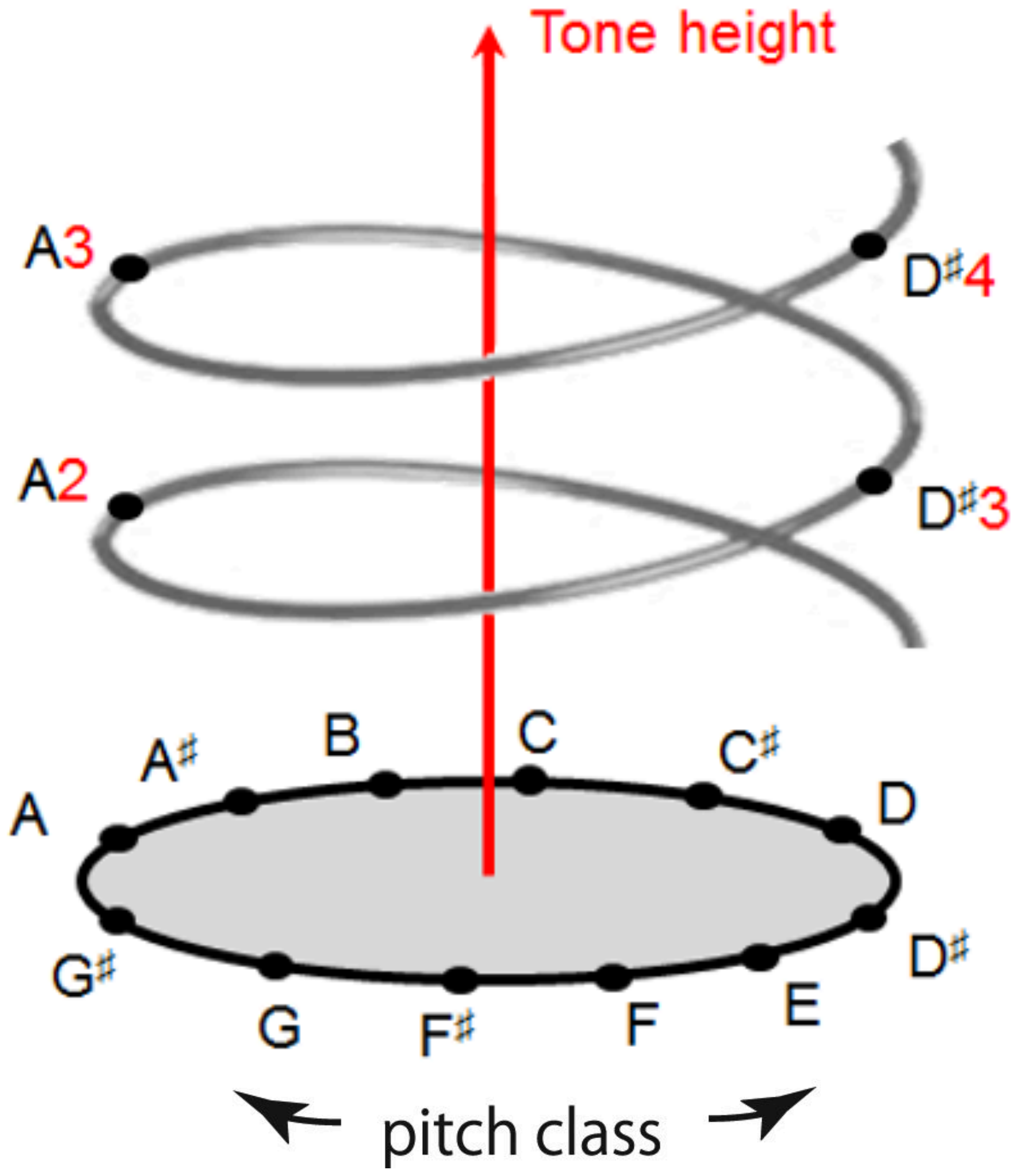
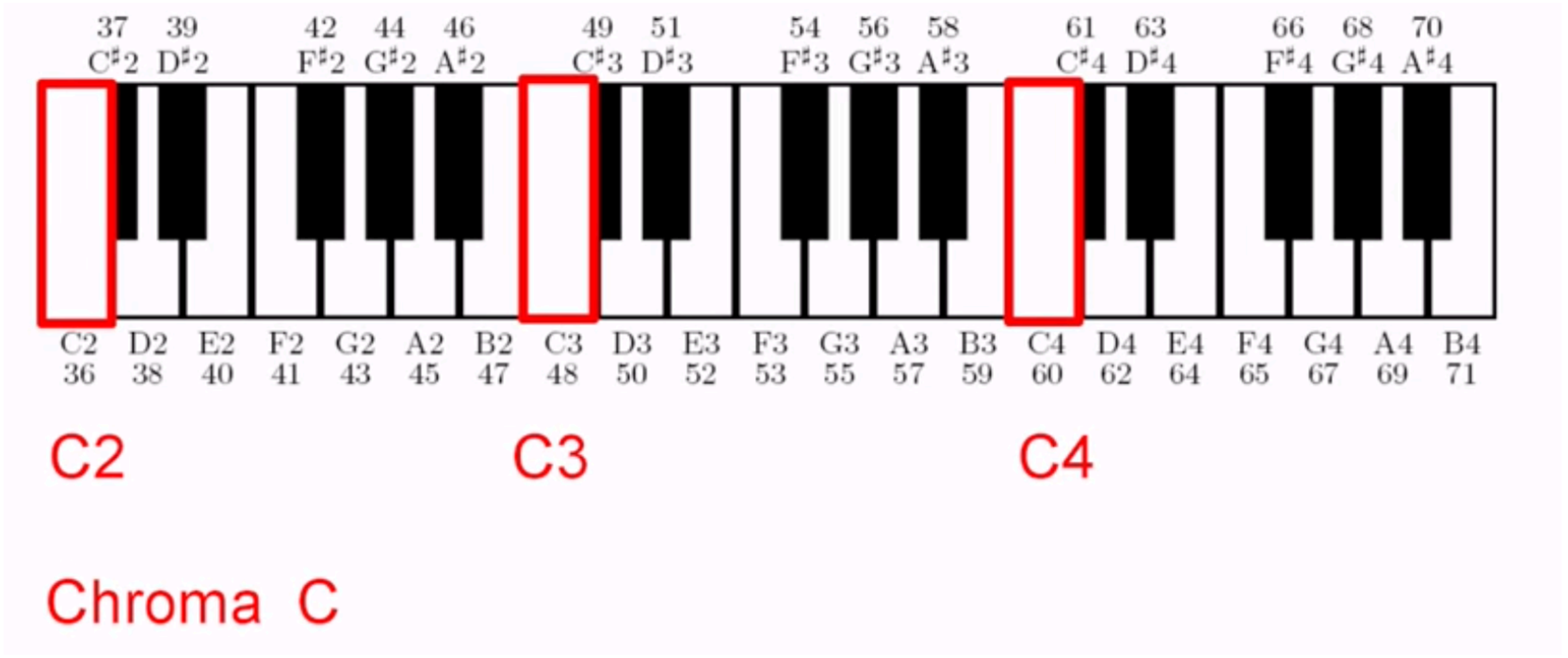
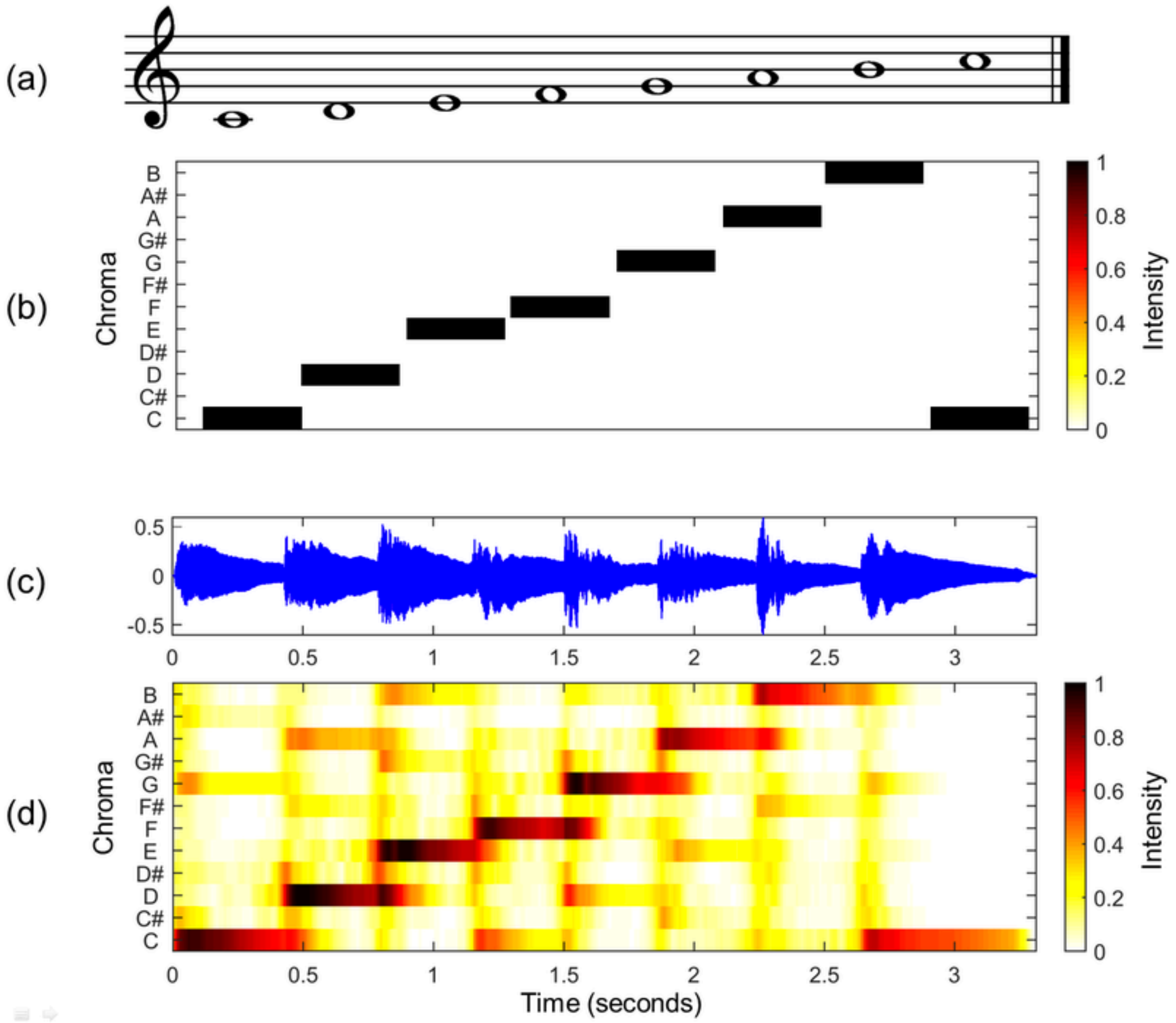


Figure 1.3 from [Müller, FMP, Springer 2015]





Chromagram

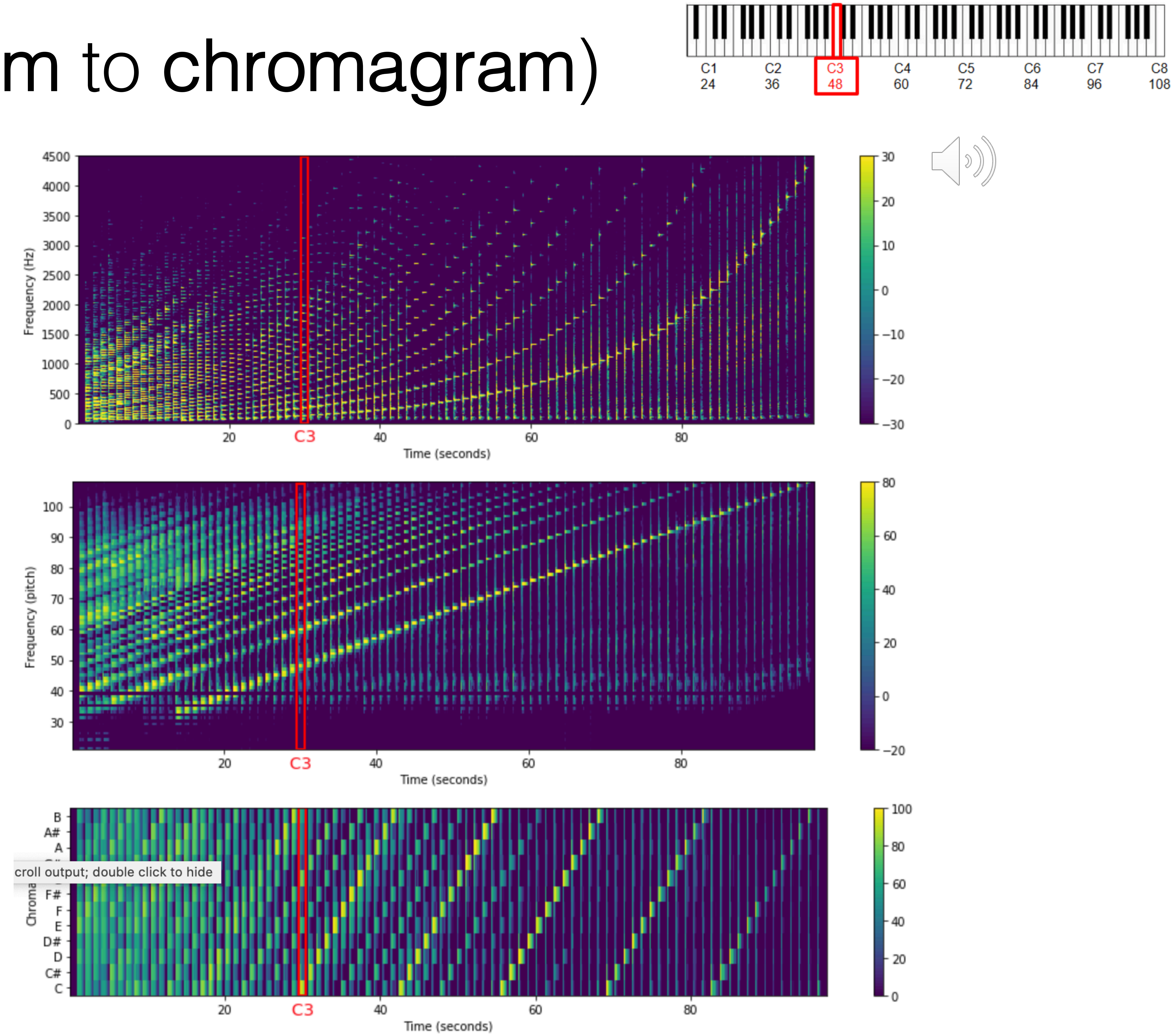


from: https://en.wikipedia.org/wiki/Chroma_feature

Chromagram

- We can collapse all of the octaves of the spectrum into a **single octave** to make the chromagram
- Notes that belong to the same pitch class (or have the same chroma value) are perceived as similar in a certain way (same color/chroma).
- In contrast, notes that belong to different pitch classes (or have different chroma values) are perceived as dissimilar (different color/chroma).
- The chromagram is highly useful **mid-level representation** of harmony in many MIR tasks (**chords, key, structure segmentation, cover song identification, music similarity, etc**.)

Chromagram (from **spectrogram** to **chromagram**)



Spectrogram

Log-Frequency Spectrogram

We obtain a log-frequency spectrogram $\mathcal{Y}_{\mathrm{LF}} : \mathbb{Z} \times [0:127]$ using a simple pooling procedure:

$$\mathcal{Y}_{\mathrm{LF}}(n,p) := \sum_{k \in P(p)} |\mathcal{X}(n,k)|^2.$$

By this procedure, the frequency axis is partitioned logarithmically and labeled linearly according to **MIDI pitches**.

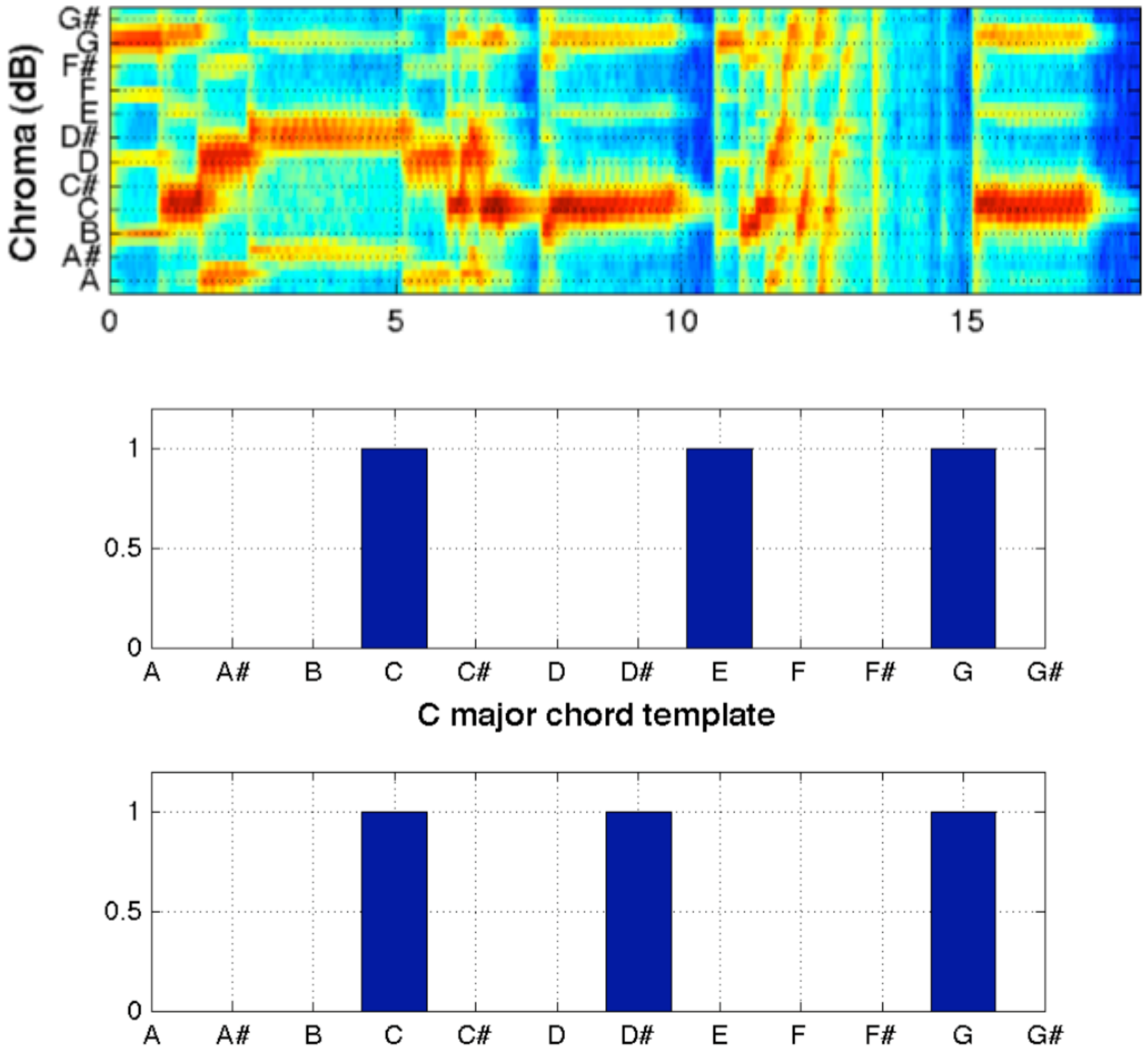
$F_{\mathrm{pitch}}(p) = 2^{(p-69)/12} \cdot 440$. (where the note A4 corresponds to MIDI note number $p = 69$)

Chromagram

NOTEBOOK: Log-Frequency Spectrogram and Chromagram
Section 3.1.1 of [Müller, FMP, Springer 2015],

Chord Recognition

- Compare chromagram to binary chord templates
- Attempt to “decode” the sequence of most likely chords
- Use model of chord transition probabilities - some chord changes are much more likely than others



C major chord template

C minor chord template

image from: http://www.nyu.edu/classes/bello/MIR_files/tonality.pdf

Chord Recognition

- There are many approaches to chord recognition
- However, there is only a moderate difference in accuracy between the simplest and most complex methods
- Beat tracking is often employed to give "beat-synchronous chromagrams"

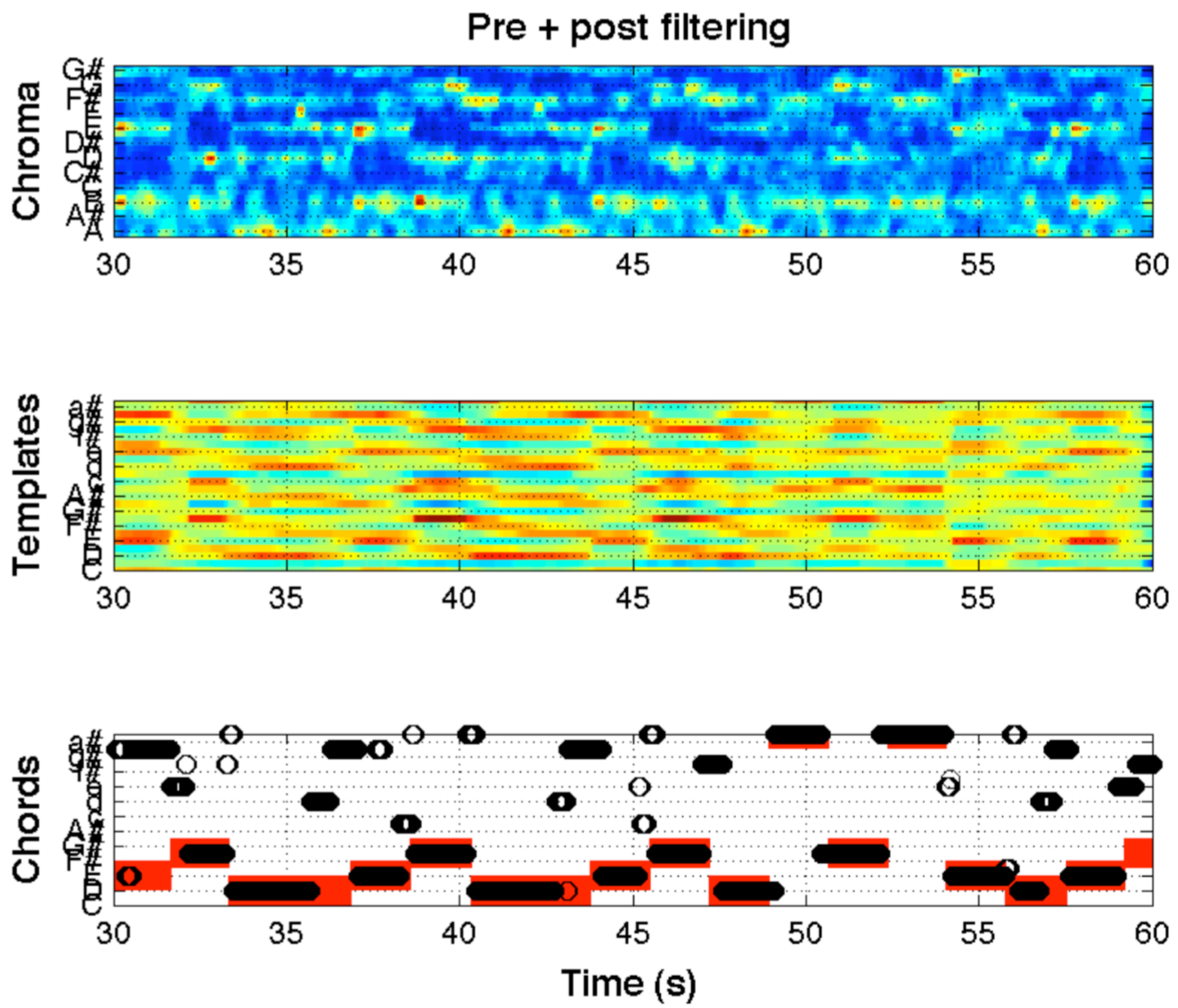


image from: http://www.nyu.edu/classes/bello/MIR_files/tonality.pdf

Key Estimation

- Find the predominant key
- Use profiles of strengths of notes in each key
- Compare to chromagram
- Extract most likely key (among 12 major and 12 minor keys)
- Can track key changes through time, but perform analysis over long windows (e.g. > 20s)

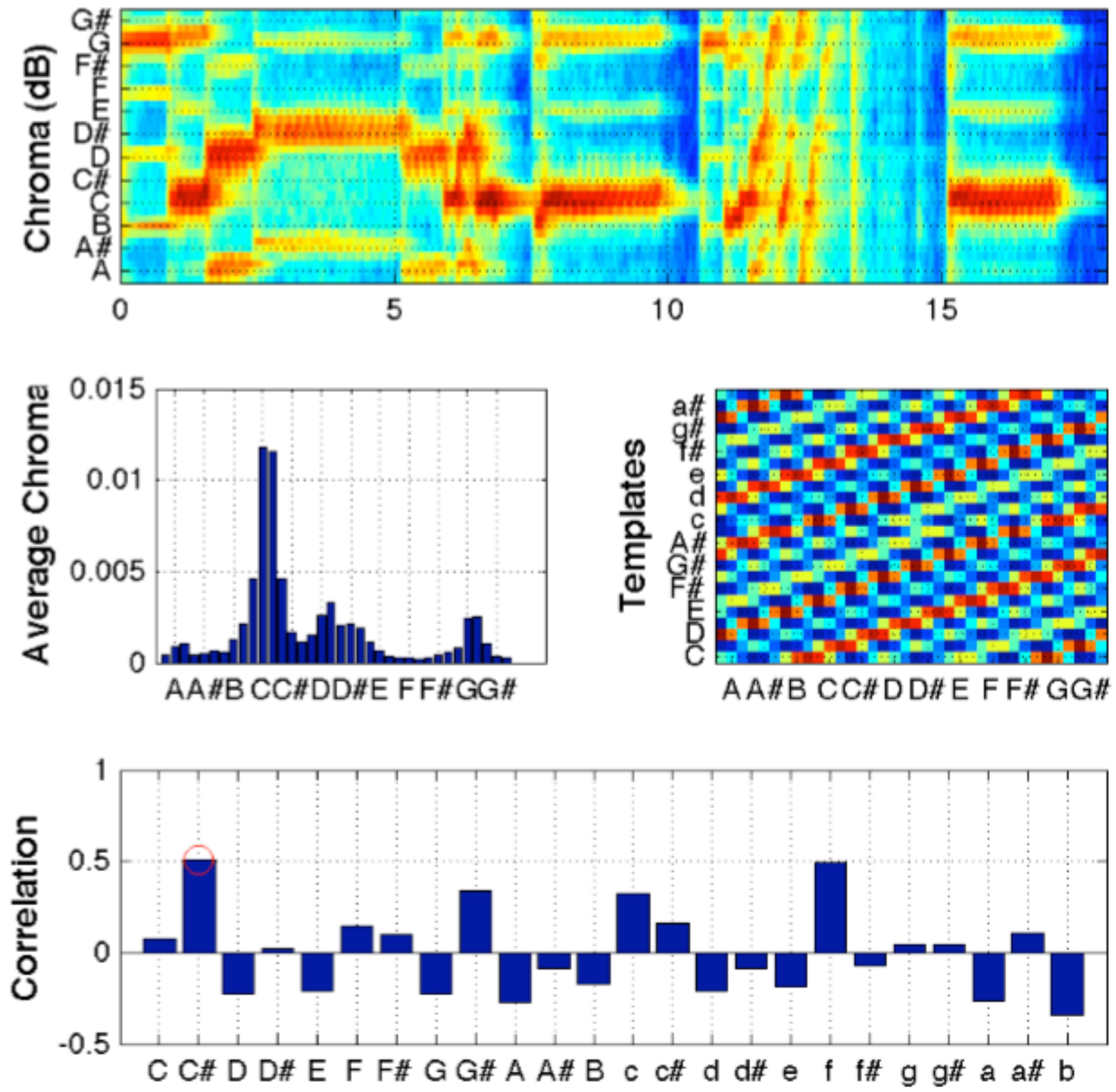


image from: http://www.nyu.edu/classes/bello/MIR_files/tonality.pdf

chroma, key and chord summary

It's difficult to obtain "clean" constant-Q and (especially) chroma representations

A lot of effort in the MIR community to filter out percussive content and explore different enhancement methods

Labelling of chords can be ambiguous and based on context

Evaluation of chord detection algorithms is highly complex

Perception of key changes is a slow process (we may only be aware of the key change some time after the score would indicate it has occurred)

Practical Class

(GO TO MOODLE)
Task1/Task2

- As soon as you're finished (both tasks), call me.

At 16:30 we'll start the Assignment 3!

4.2. Pitch

.