Lecture 7

Pitch Analysis

What is pitch?

- (ANSI) That attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch **depends mainly on the frequency** content of the sound stimulus, but **also depends on the sound pressure and waveform** of the stimulus.

- (operational) A sound has a certain pitch if it can be reliably matched to a sine tone of a given frequency at 40 dB SPL.

What is pitch?

- A perceptual attribute, so subjective
- Only defined for (quasi) harmonic sounds
  - Harmonic sounds are periodic, and the period is 1/F0.
- Can be reliably matched to fundamental frequency (F0)
  - In audio signal processing, people do not often discriminate pitch from F0
- F0 is a physical attribute, so objective

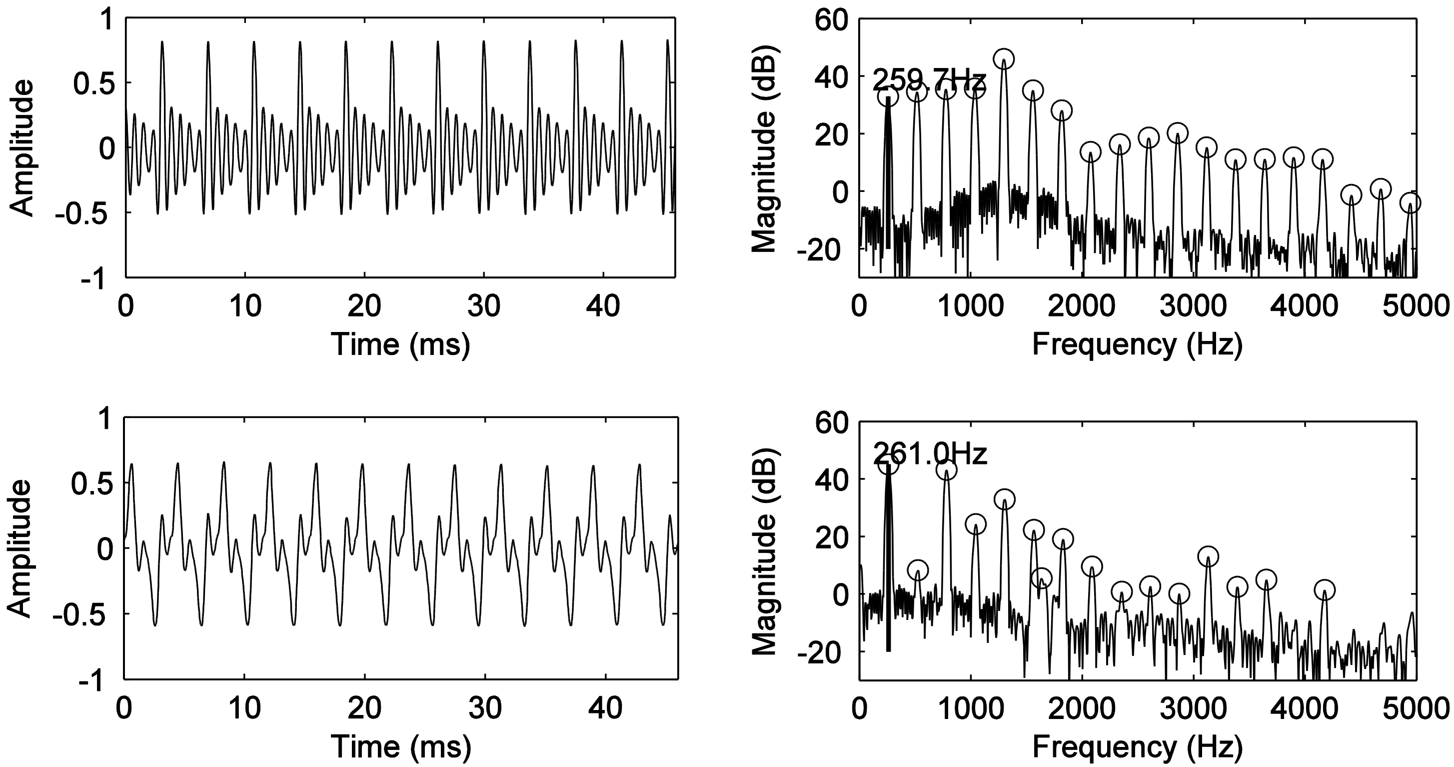
Harmonic Sound

- A complex sound with strong sinusoidal components at integer multiples of a fundamental frequency. These components are called **harmonics** or **overtones**.

- Sine waves and harmonic sounds are the sounds that may give a perception of "pitch".

Classify Sounds by Harmonicity

- Sine wave
- Strongly harmonic

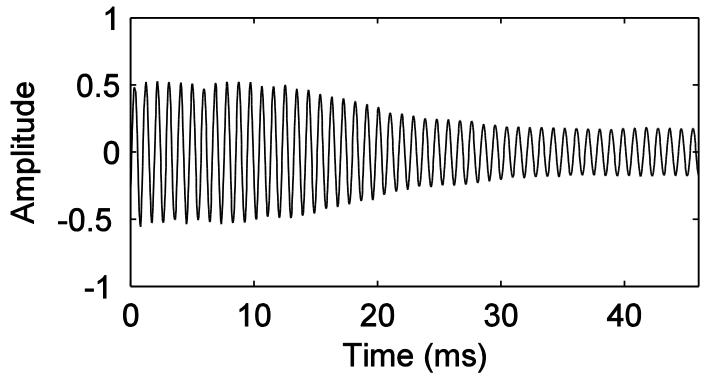


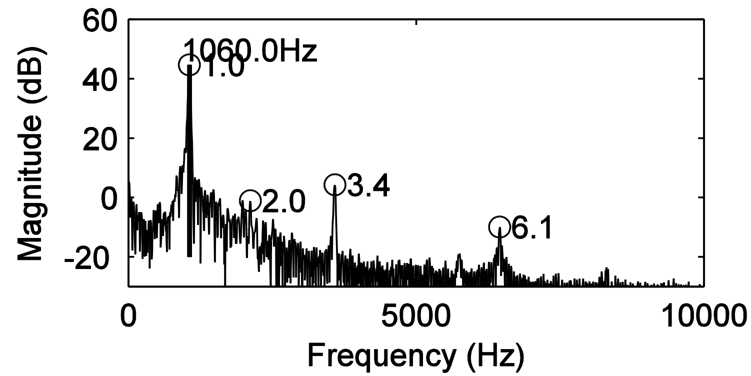
Oboe

Clarinet

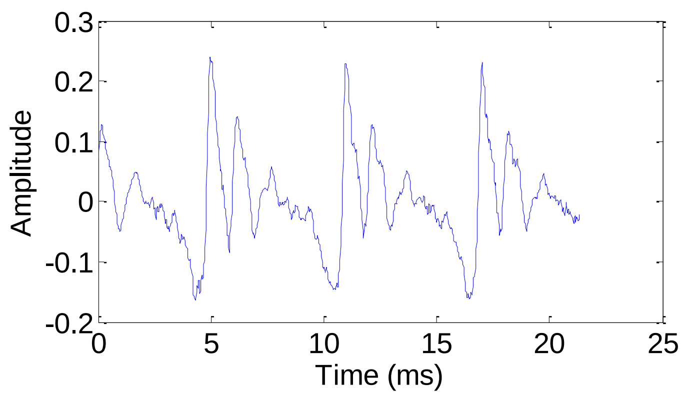
Classify Sounds by Harmonicity

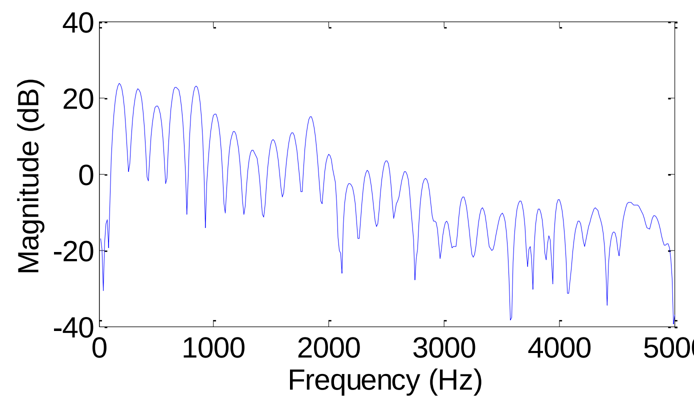
- Somewhat harmonic (quasi-harmonic)





Marimba

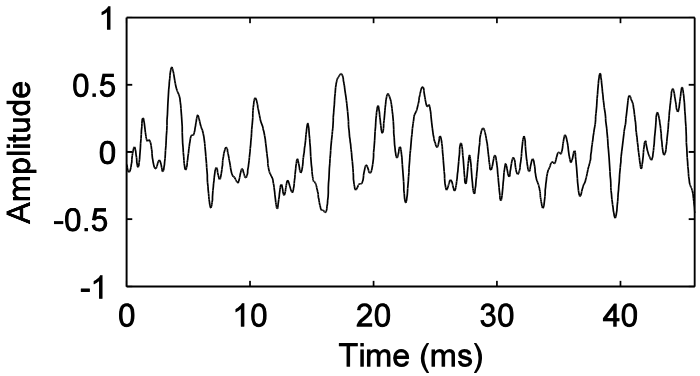


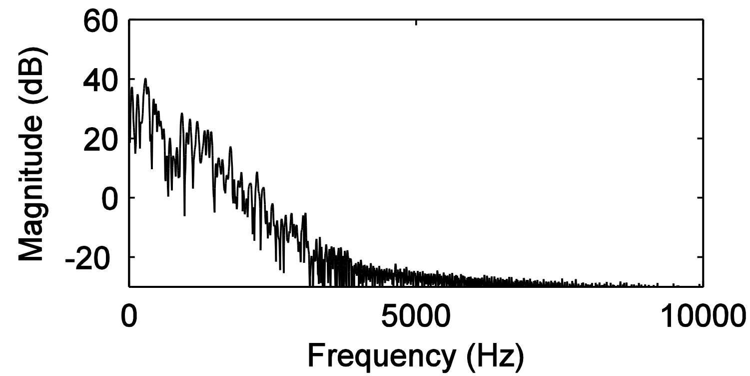


Human voice

Classify Sounds by Harmonicity

- Inharmonic



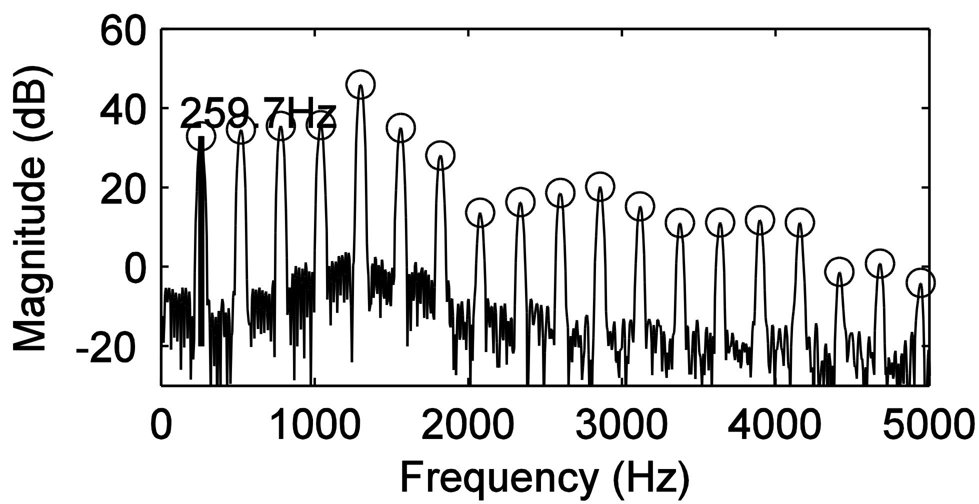


Gong

Sounds	Instrument family	Instruments
	Woodwind	Piccolo, flute, oboe, clarinet, bassoon, saxophone
	Brass	Trumpet, horn, euphonium, trombone, tuba
Harmonic	Arco string	Violin, viola, cello, double bass
	Pluck string	Piano, guitar, harp, celesta
	Vocal	Voiced phonemes
Quasi-harmonic	Pitched percussive	Timpani, marimba, vibraphone, xylophone
Inharmonic	Non-pitched percussive	Drums, cymbal, gong, tambourine

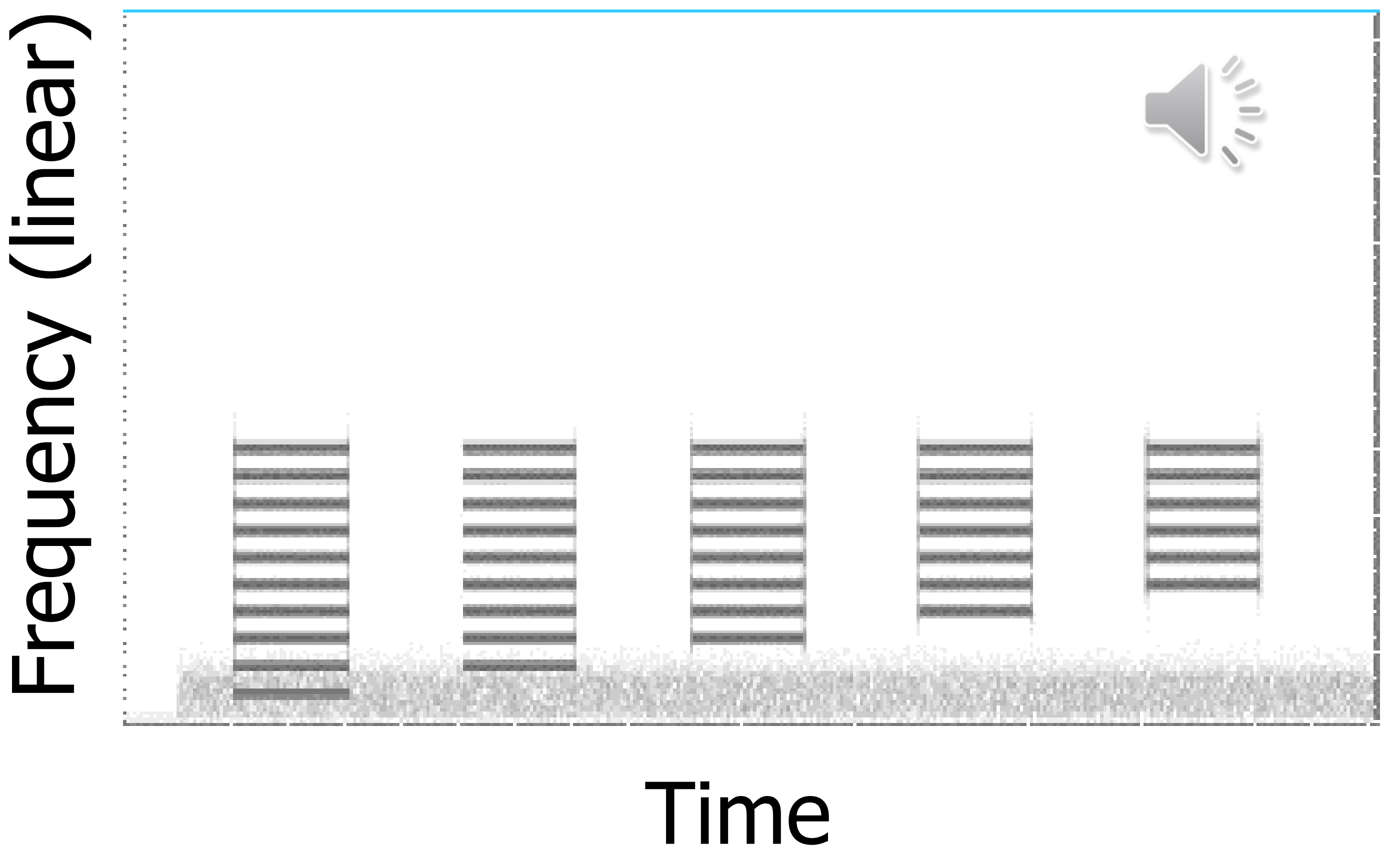
How do we perceive pitch?

- Complex tones
  - Strongest frequency?
  - Lowest frequency?
  - Greatest common divisor of the harmonics?



Oboe

The Missing Fundamental

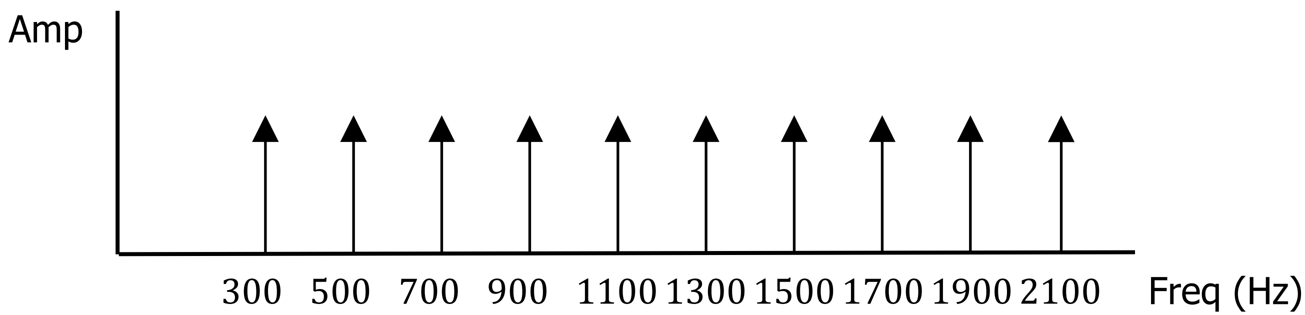


Another Example

300:200:2100 Hz

200Hz

300Hz



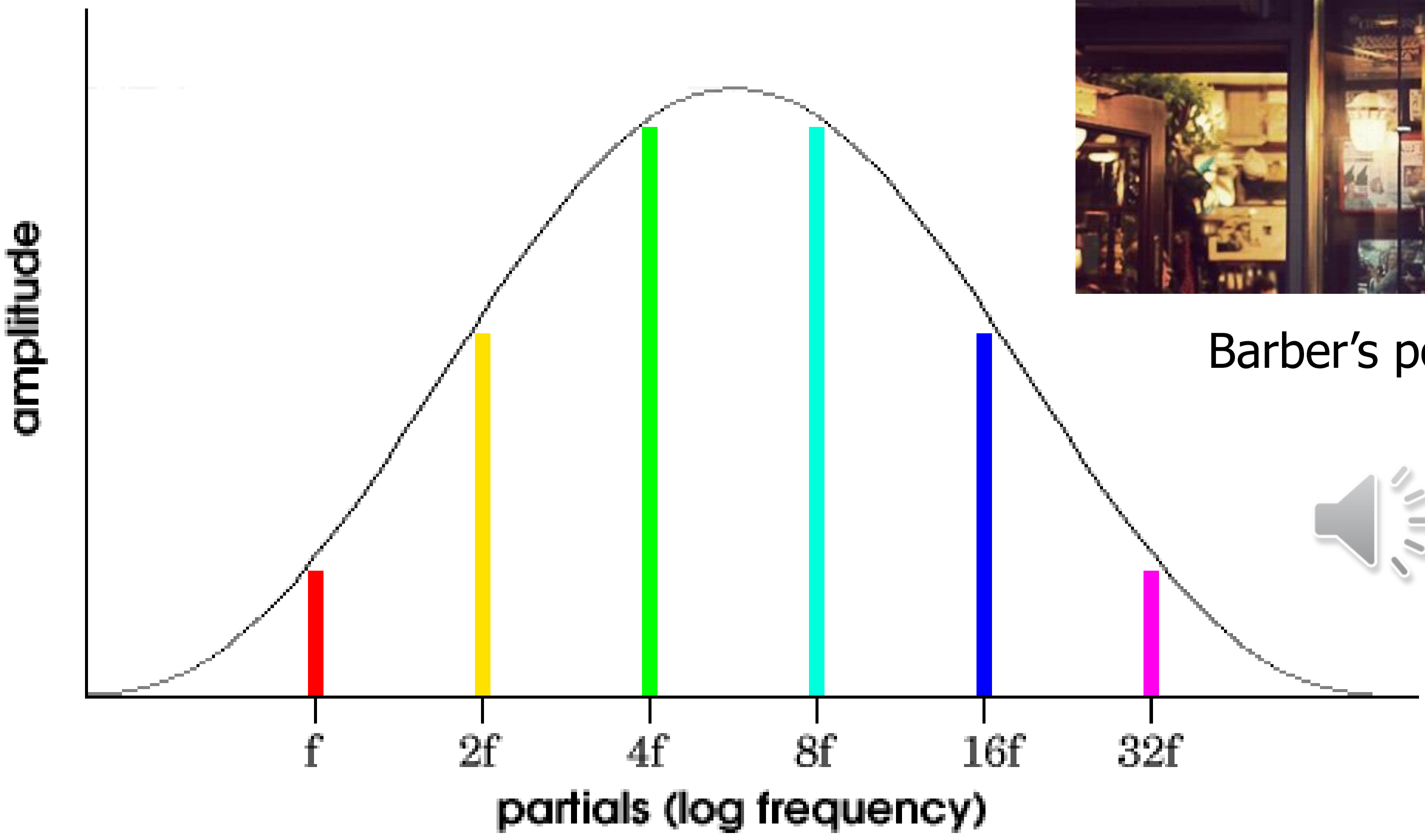
Pitch perception is affected by

- The loudest frequency component

- The greatest common divisor of partials

- The regular frequency space between partials

Shepard Tones



Barber's pole

Continuous Risset scale

Pitch Detection

- Well defined for harmonic and quasi-harmonic sounds
- Estimate the fundamental frequency in each frame of the signal

- Quick facts
  - Human speech: from 40 Hz for low-pitched male to 600 Hz for children or high-pitched female
  - Piano: 27 Hz – 4,186 Hz
  - Human hearing range: 20 Hz – 20,000 Hz

Why is pitch detection important?

- Harmonic sounds are ubiquitous
  - Music, speech, bird singing
- Pitch (F0) is an important attribute of harmonic sounds, and it relates to other properties
  - Music melody → key, scale (e.g., chromatic, diatonic, pentatonic), style, emotion, etc.
  - Speech intonation → word disambiguation (for tonal language), statement/question, emotion, etc.

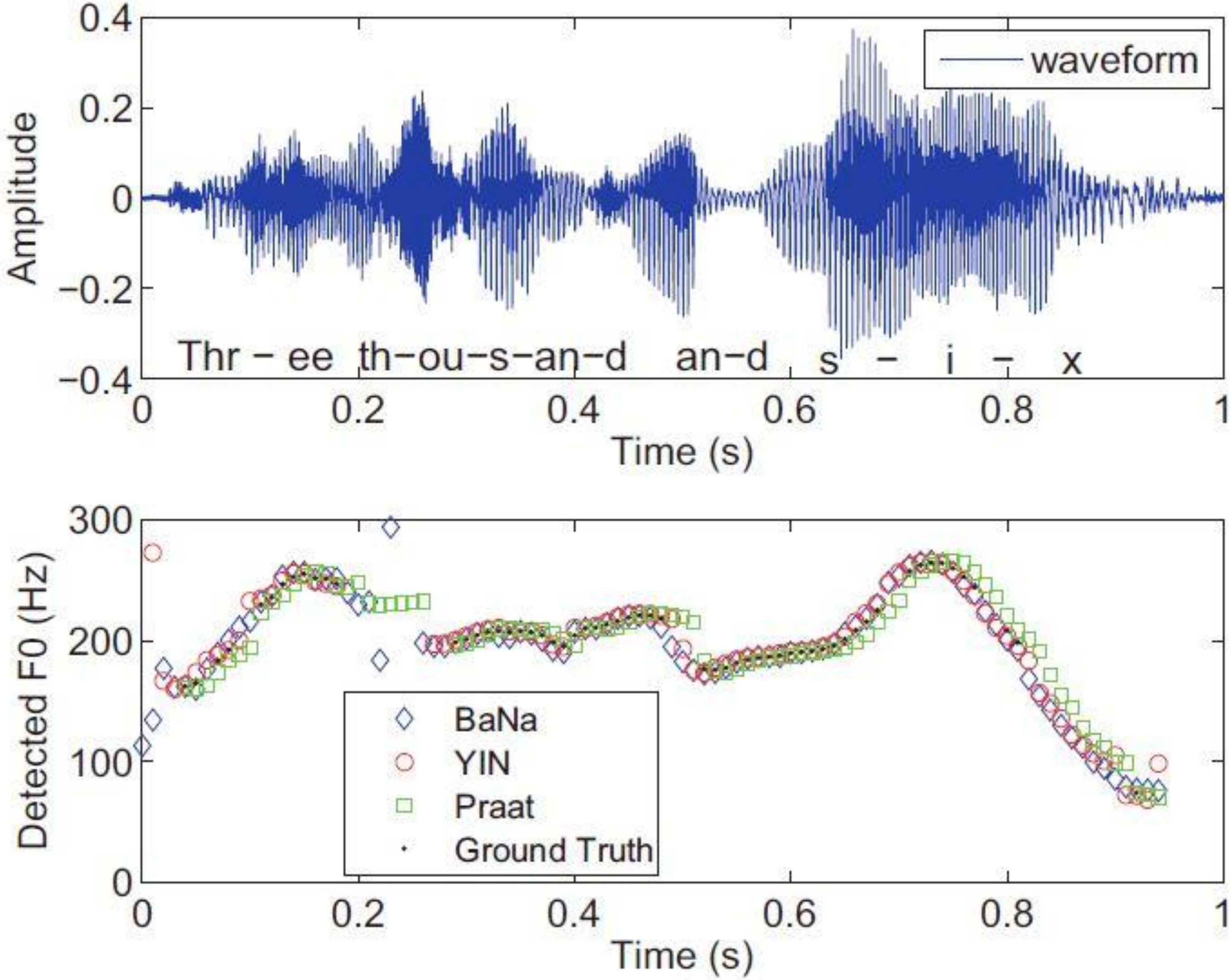
What scales are used?

What emotion?

General Process of Pitch Detection

- Segment audio into time frames
  - Pitch changes over time
- Detect pitch (if any) in each frame
  - Need to detect whether the frame contains pitch or not
- Post processing to consider context info
  - Pitch contours are often continuous

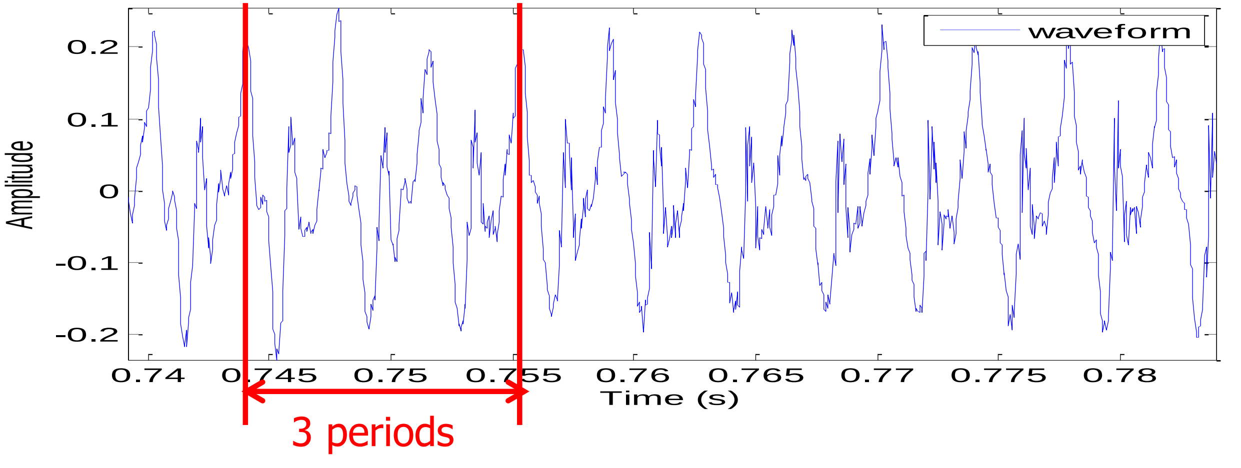
An Example



0.4
0.2
0
−0.2
−0.4
Amplitude
waveform
Thr − ee th−ou−s−an−d an−d s − i − x
0
0.2
0.4
0.6
0.8
1
Time (s)
300
200
100
0
Detected F0 (Hz)
BaNa
YIN
Praat
Ground Truth
0
0.2
0.4
0.6
0.8
1
Time (s)

How long should the frame be?

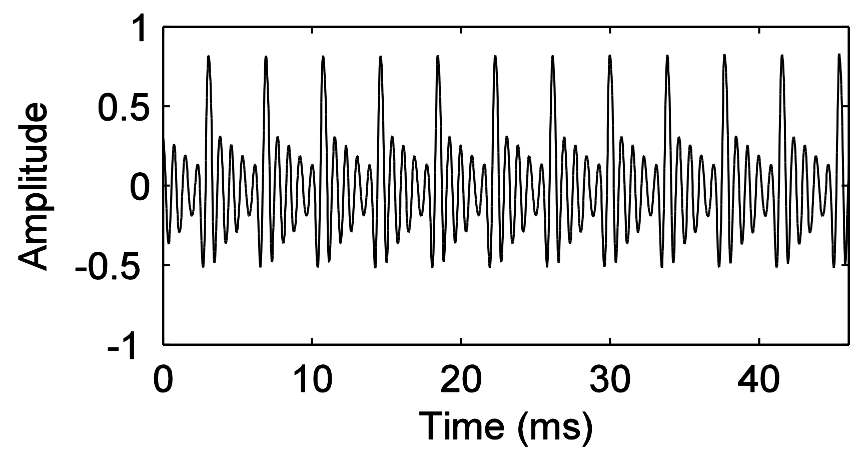
- Too long:
  - Contains multiple pitches (low time resolution)
- Too short
  - Can't obtain reliable detection (low freq resolution)
  - Should be longer than 3 periods of the signal



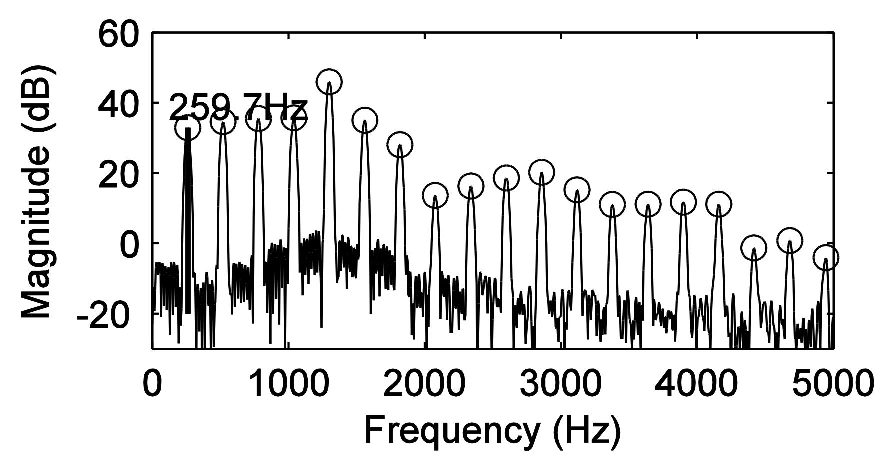
  - For speech or music, how long should the frame be?

Pitch-related Properties

- Time domain signal is periodic.
  - F0 = 1/period



- Spectral peaks have harmonic relations.
  - F0 is the greatest common divisor.



- Spectral peaks are equally spaced.
  - F0 is the frequency gap.

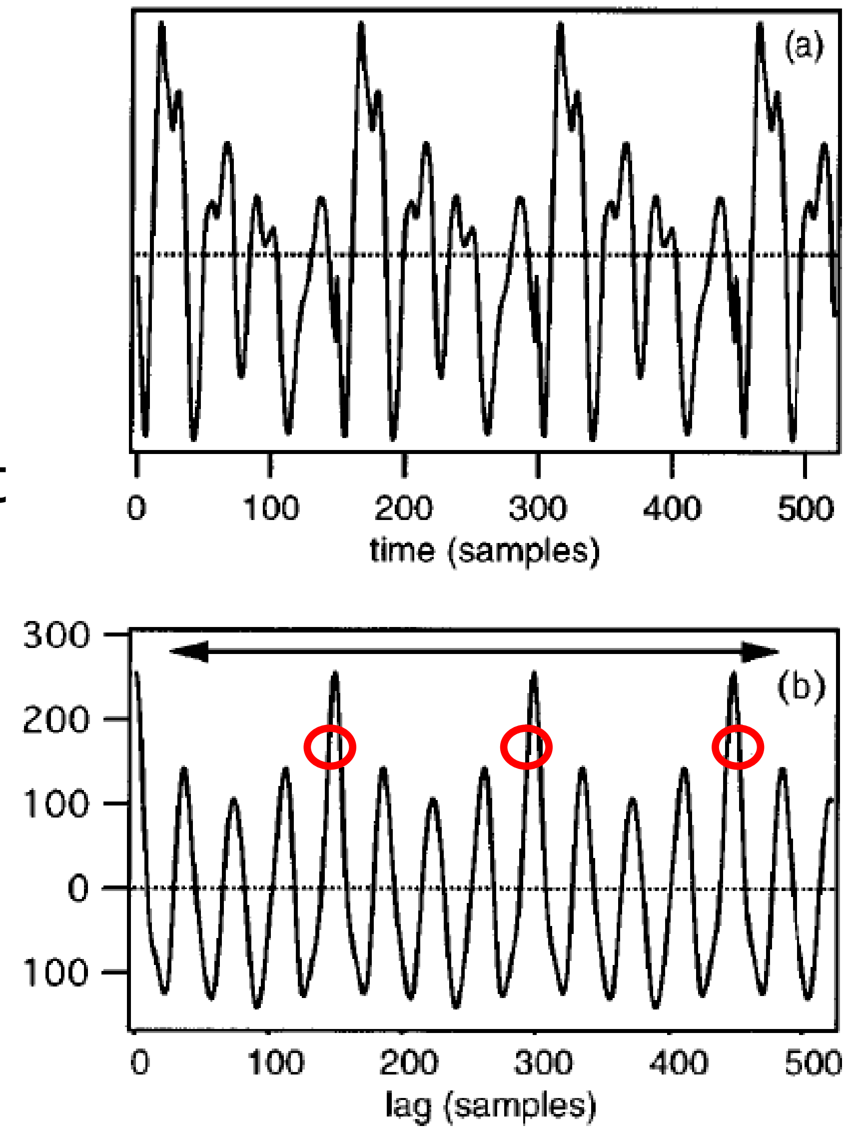
Pitch Detection Methods

- Time domain signal is periodic.
  - F0 = 1/period
- Time domain
  - Detect period
- Spectral peaks have harmonic relations.
  - F0 is the greatest common divisor.
- Frequency domain
  - Detect the divisor
- Spectral peaks are equally spaced.
  - F0 is the frequency gap.
- Cepstrum domain
  - Detect the gap

Pitch detection in time domain

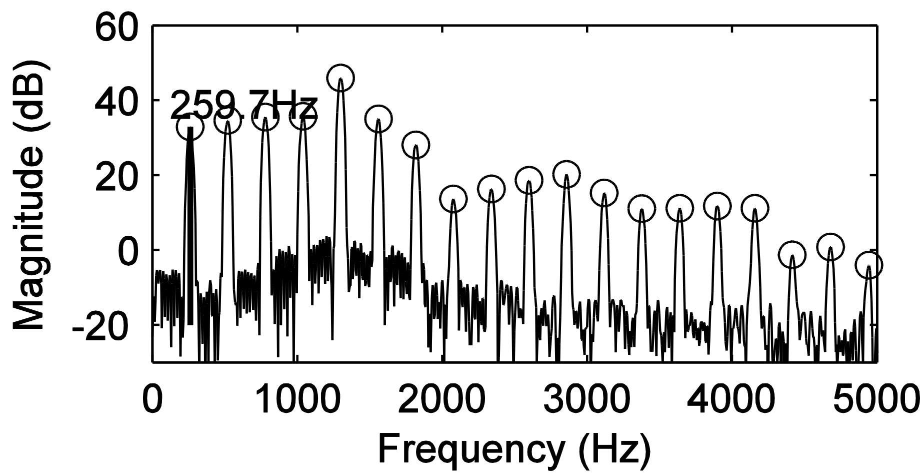
- Autocorrelation
  - Basis: the time-domain signal is periodic
  - A periodic signal correlates strongly with itself when offset by the fundamental period.
  - Autocorrelation shows peaks at multiples of pitch period.

$$r_t(\tau) = \sum_{j=t+1}^{t+W} x_j x_{j+\tau}$$



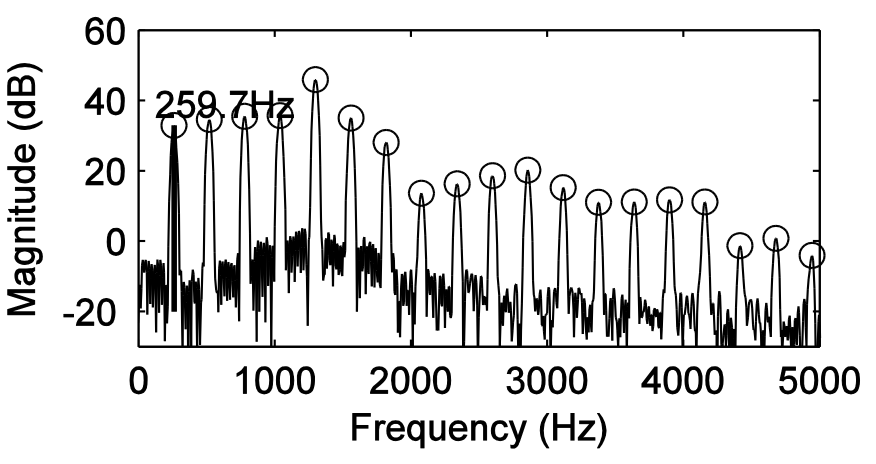
Pitch detection in frequency domain

- Calculate the magnitude spectrum
- For each pitch hypothesis, calculate salience by
  - Counting the number of peaks located at its harmonic positions, or
  - Summing spectral energy at harmonic positions, or
  - ……
- Choose the hypothesis with the highest salience



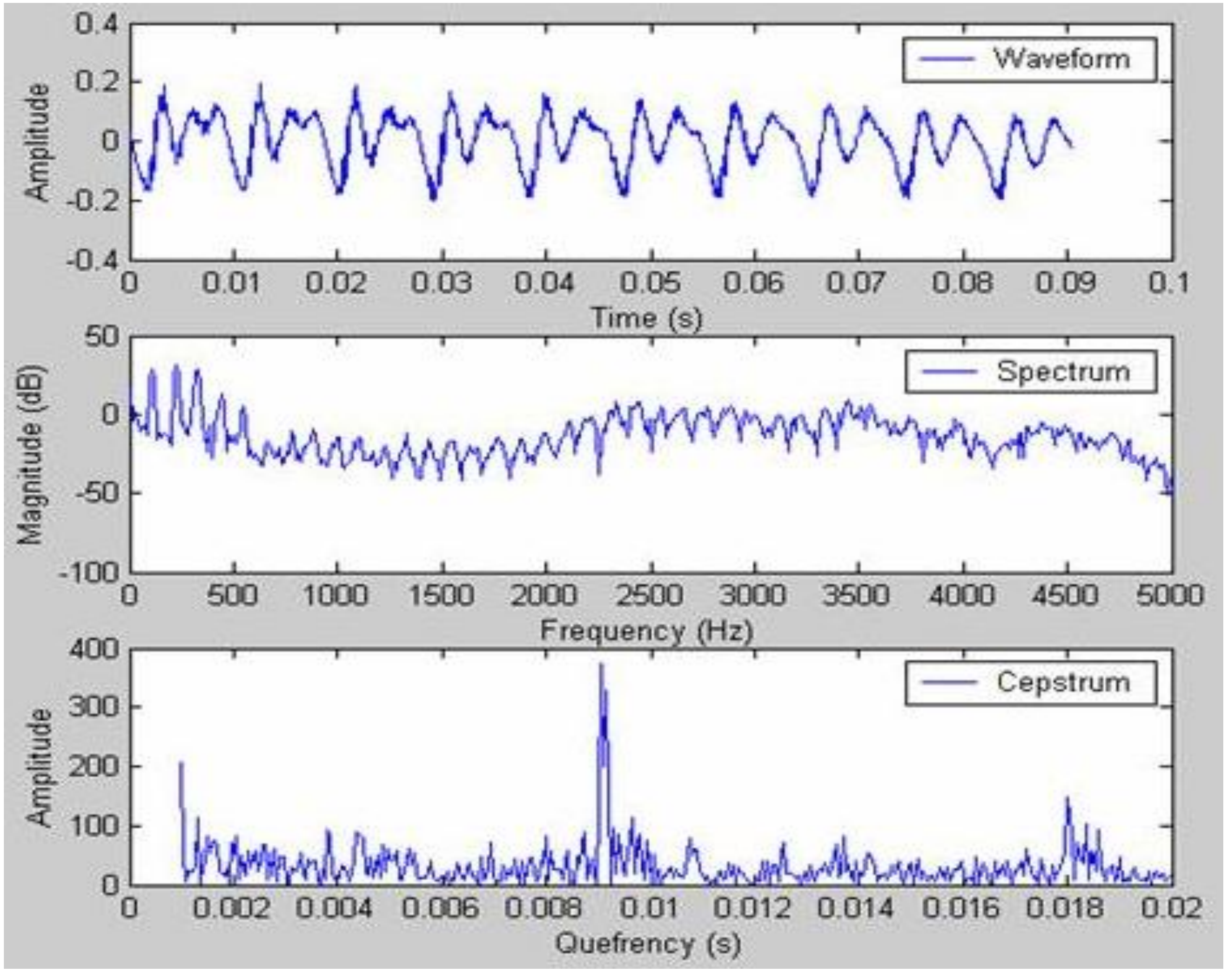
Pitch detection in cepstrum domain

- Idea: find the frequency gap between adjacent spectral peaks
  - The log-amplitude spectrum looks pretty periodic
  - The gap can be viewed as the period of the spectrum
  - How to find the period then?
  - Cepstrum's idea: Fourier transform!



Pitch detection in cepstrum domain

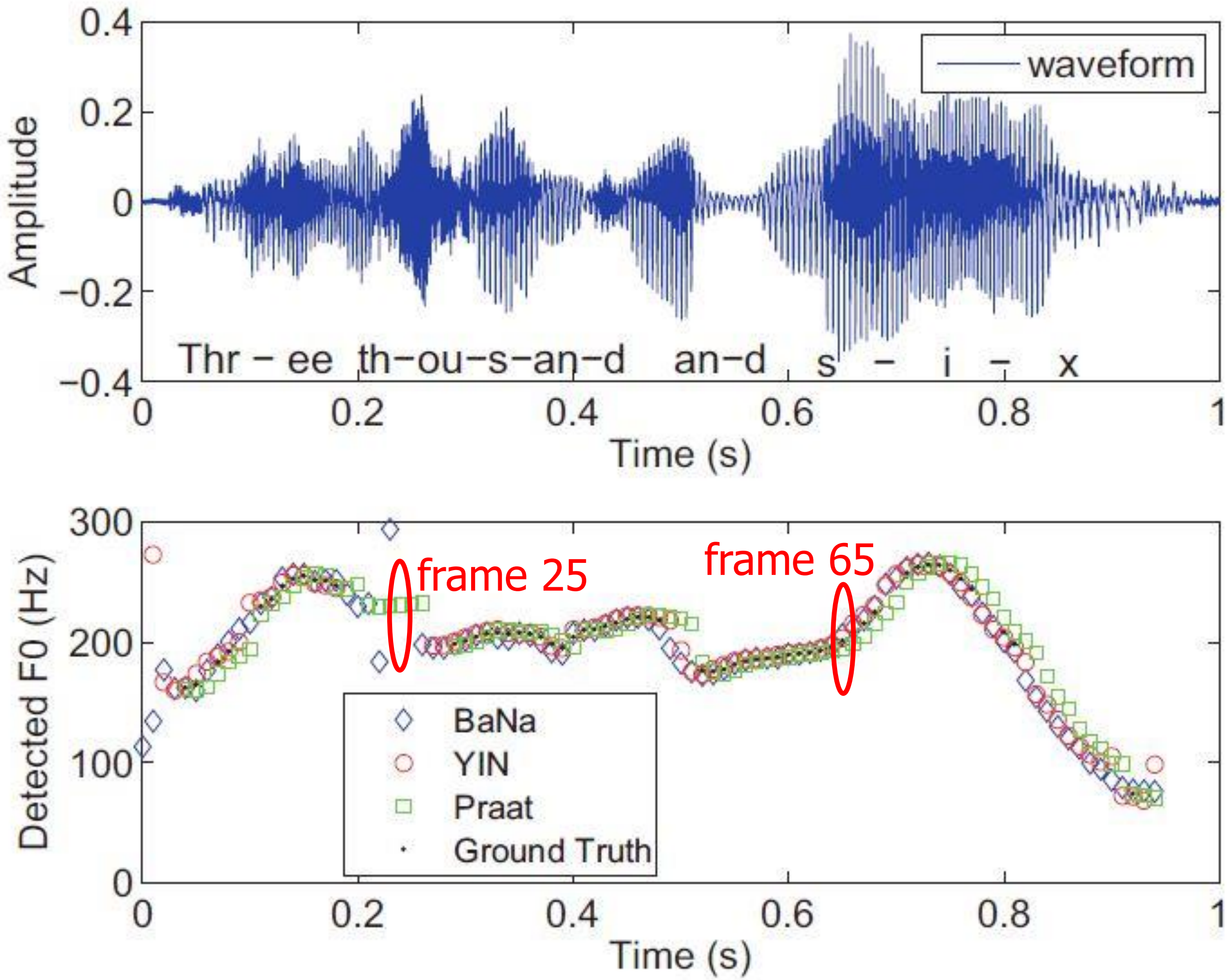
- Cepstrum = |IFT{log|FT(X)|}|



How to evaluate pitch detection?

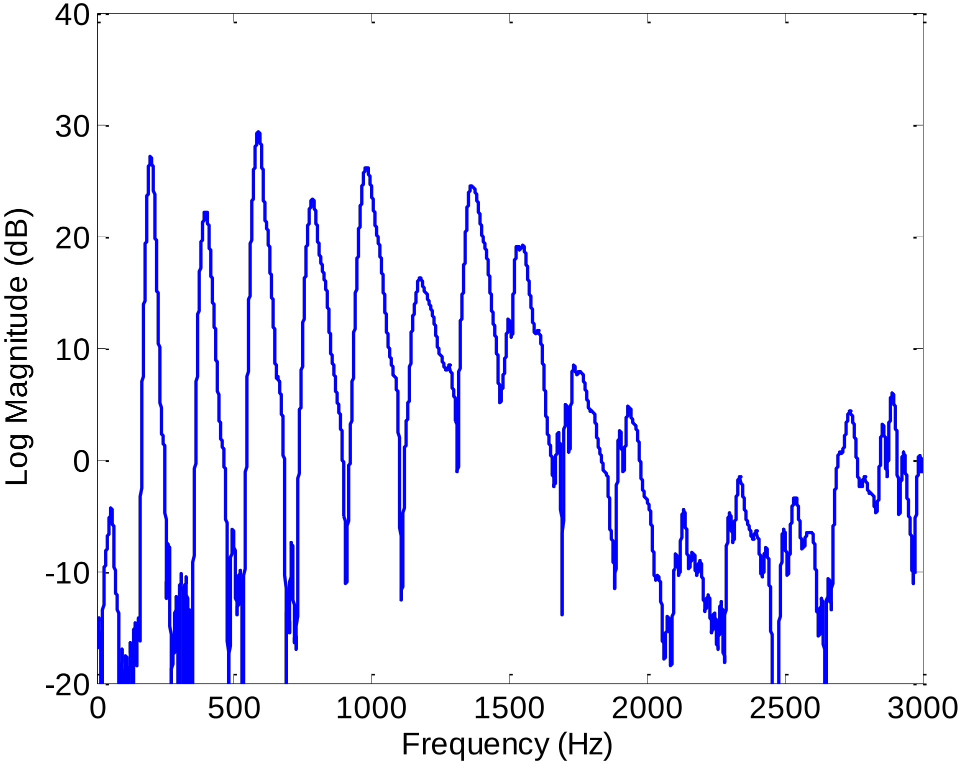
- Choose some recordings (speech, music)
- Get ground-truth
  - Listen to the signal and inspect the spectrum to manually annotate (time consuming!)
  - Automatic annotation using simultaneously recorded laryngograph signals for speech (not quite reliable!)
- Pitched/non-pitched classification error
- Calculate the difference between estimated pitch and ground-truth
  - Threshold for speech: 10% or 20% in Hz
  - Threshold for music: 1 quarter-tone (about 3% in Hz)

Different Methods vs. Ground-truth



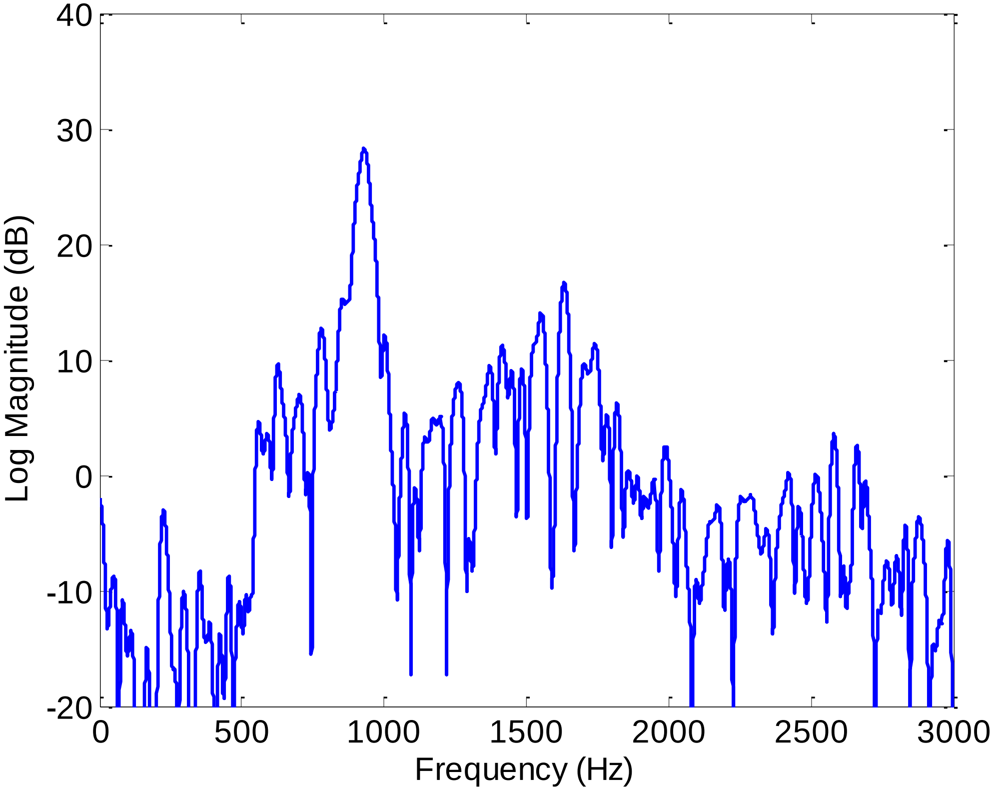
Frame 65 – Pitched (Voiced)

- Has clear harmonic patterns
- Different methods give close results, and consistent to the ground-truth 196 Hz.



Frame 25 – Non-pitched (Unvoiced)

- No clear harmonic patterns
- Different methods give inconsistent results.



Pitch Detection with Noise

- Can we still hear pitch if there is some background noise, say in a restaurant?

Violin + babble noise

- Will pitch detection algorithms still work?
- Which domain is less sensitive to what kind of noise?
- How to improve pitch detection in noisy environments?

Summary

- Pitch detection is important for many tasks
  - Time domain: find the period of waveform
  - Frequency domain: find the divisor of peaks
  - Cepstrum domain: find the frequency gap between spectral peaks

- Single pitch detection is mature in noiseless conditions.

- Pitch detection in noisy environments (also called robust pitch detection, noise-resilient pitch detection) is an active research topic.
  - BaNa [Yang et al., 2014]; PEFAC [Gonzales & Brookes, 2014];

- Multi-pitch Estimation is extremely challenging!