University of Victoria Computer Science

CSC475 Music Information Retrieval

Monophonic pitch extraction

George Tzanetakis

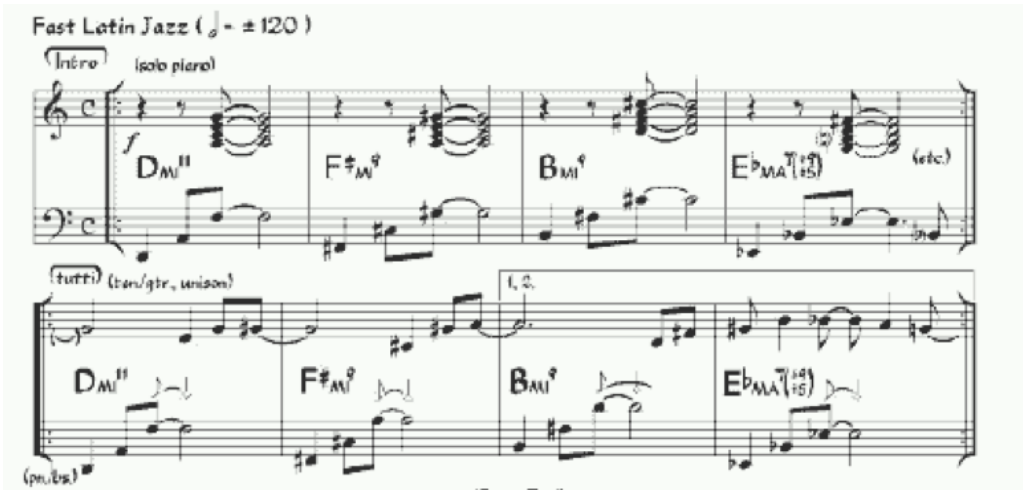
University of Victoria

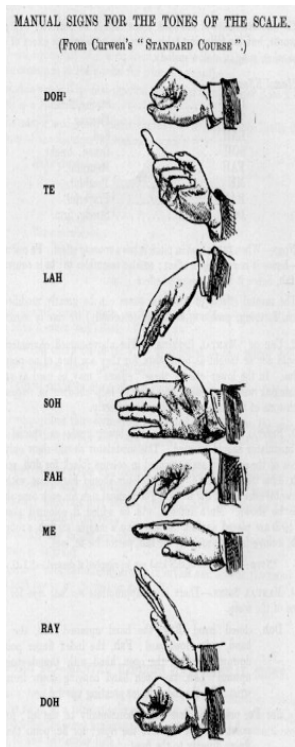
2014

Table of Contents I

Music Notation

Music notation systems typically encode information about discrete musical pitch (notes on a piano) and timing.





Terminology

The term pitch is used in different ways in the literature which can result in some confusion.

Perceptual Pitch: is a perceived quality of sound that can be ordered from "low" to "high".

Musical Pitch: refers to a discrete finite set of perceived pitches that are played on musical instruments

Measured Pitch: is a calculated quantity of a sound using an algorithm that tries to match the perceived pitch.

Monophonic: refers to a piece of music in which a single sound source (instrument or voice) is playing and only one pitch is heard at any particular time instance.

Table of Contents I

Psychoacoustics

Definition

The scientific study of sound perception.

Frequently testing the limits of perception:

- Frequency range 20Hz-20000Hz
- Intensity (0dB-120dB)
- Masking
- Missing fundamental (presence of harmonics at integer multiples of fundamental give the impression of "missing" pitch)

Origins of Psychoacoustics

Pythagoras of Samos established a connection between perception (music intervals) and physical measurable quantities (string lengths) using the monochord.

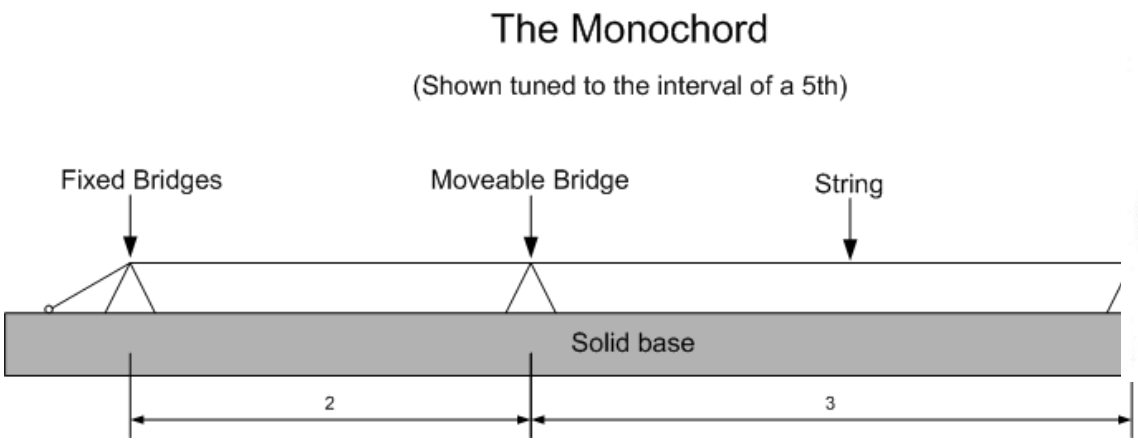
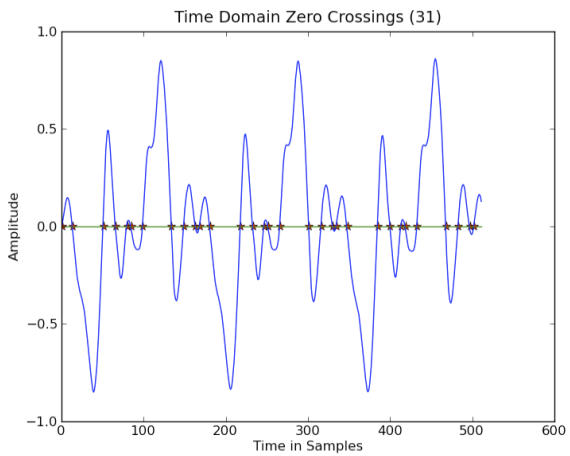


Table of Contents I

Pitch Detection

Pitch is a **PERCEPTUAL** attribute correlated but not equivalent to fundamental frequency. Simple pitch detection algorithms most deal with fundamental frequency estimation but more sophisticated ones take into account knowledge about the human auditory system.

- Time Domain
- Frequency Domain
- Perceptual



Time-domain Zerocrossings

Zero-crossings are sensitive to noise so frequency low-pass filtering is utilized.

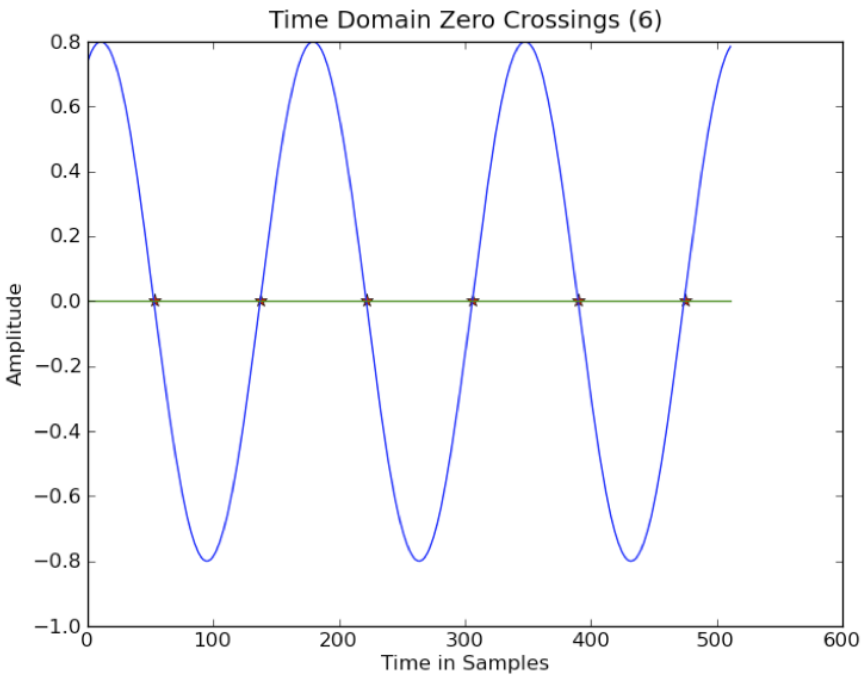


Figure: C4 Sine [Sound]

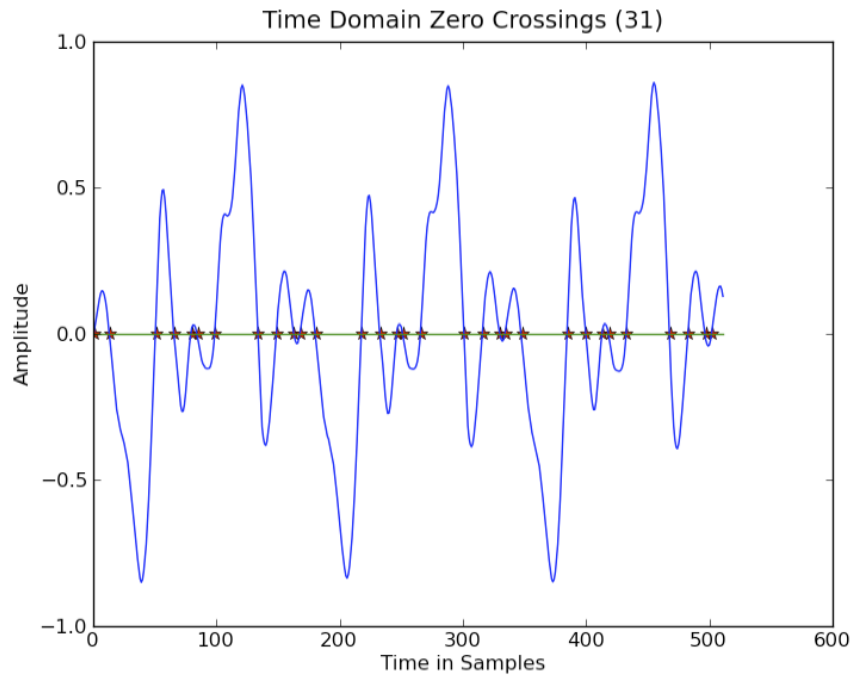


Figure: C4 Clarient [Sound]

AutoCorrelation

In autocorrelation the signal is delayed and multiplied with itself for different time lags l. The autocorrelation functions has peaks at the lags in which the signal is self-similar.

Definition

$$r_x[l] = \sum_{n=0}^{N-1} x[n]x[n+l] \quad l = 0, 1, \ldots, L-1$$

Efficient Computation

$$\begin{aligned} X[f] &= DFT\{X(t)\} \\ S[f] &= X[f]X^*[f] \\ R[l] &= DFT^{-1}\{S[f]\} \end{aligned}$$

Autocorrelation examples

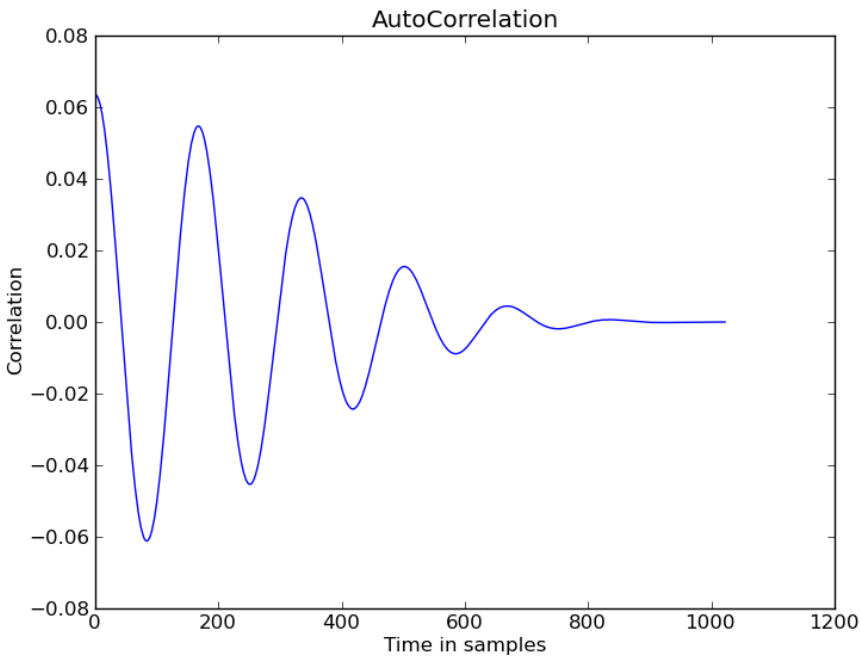


Figure: C4 Sine

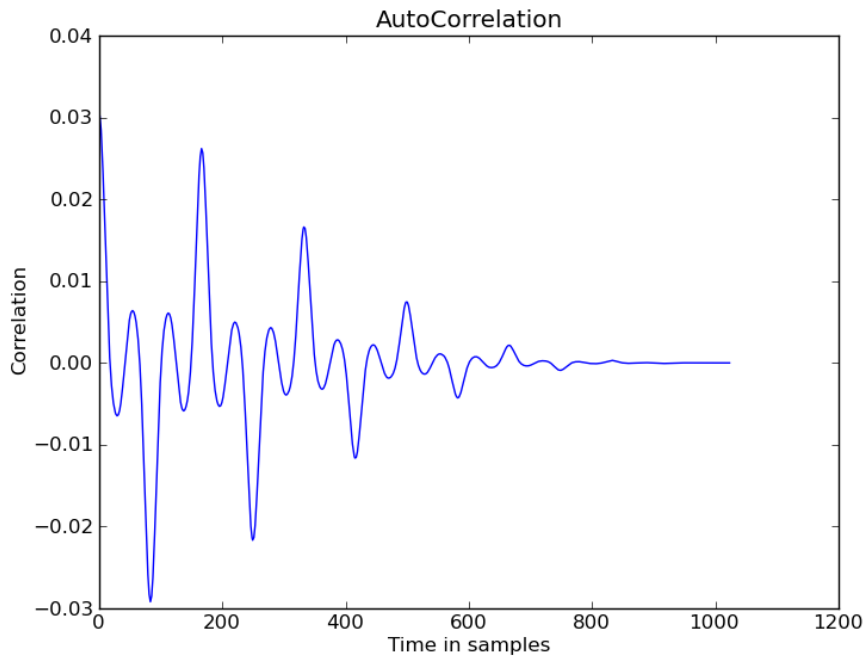


Figure: C4 Clarinet Note

Average Magnitude Difference Function

The average magnitude difference function also shifts the signal but instead of multiplication uses subtraction to detect periodicities as nulls. No multiplications make it efficient for DSP chips and real-time processing.

Definition

$$AMDF(m) = \sum_{n=0}^{N-1} |x[n] - x[n+m]|^k$$

AMDF Examples

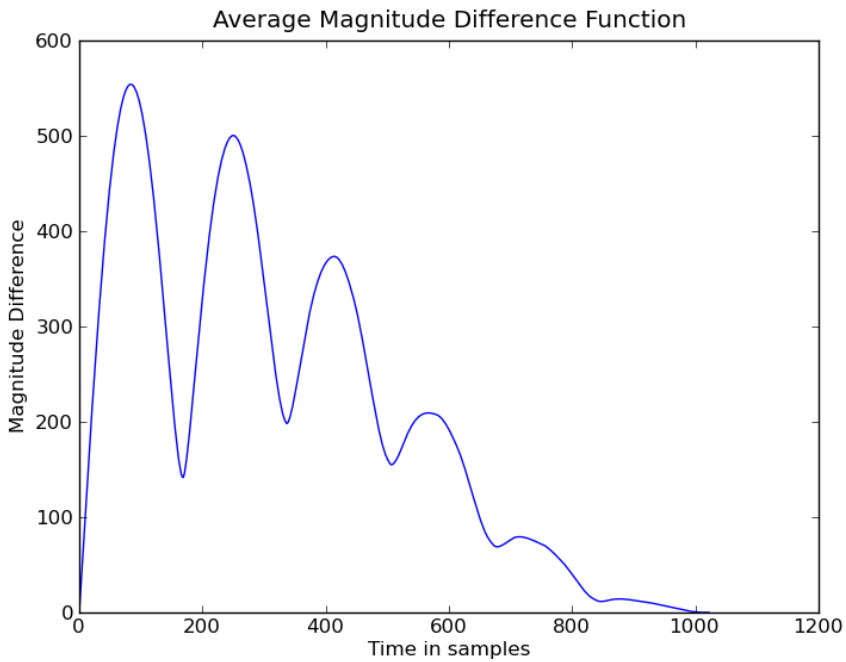


Figure: C4 Sine

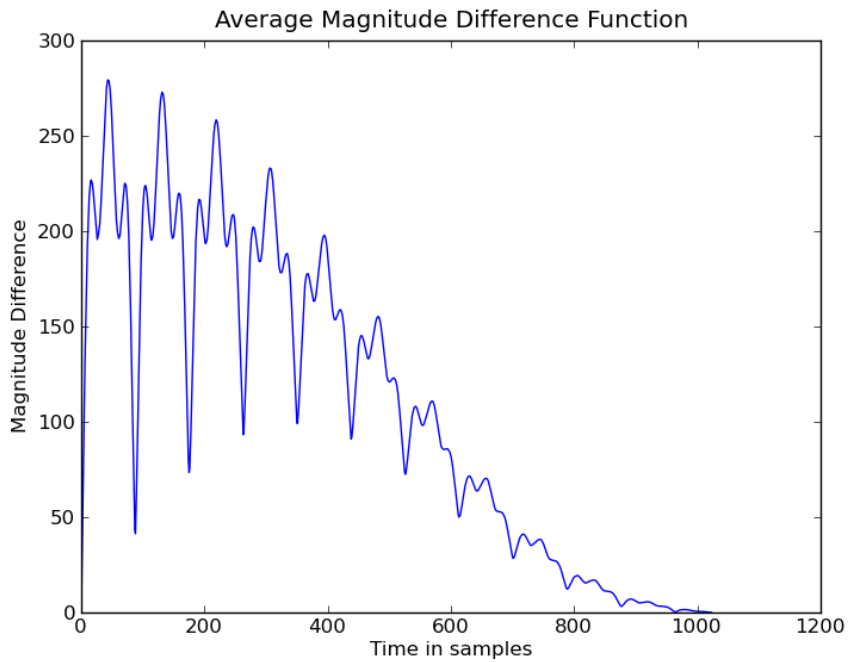


Figure: C4 Clarinet Note

Frequency Domain Pitch Detection

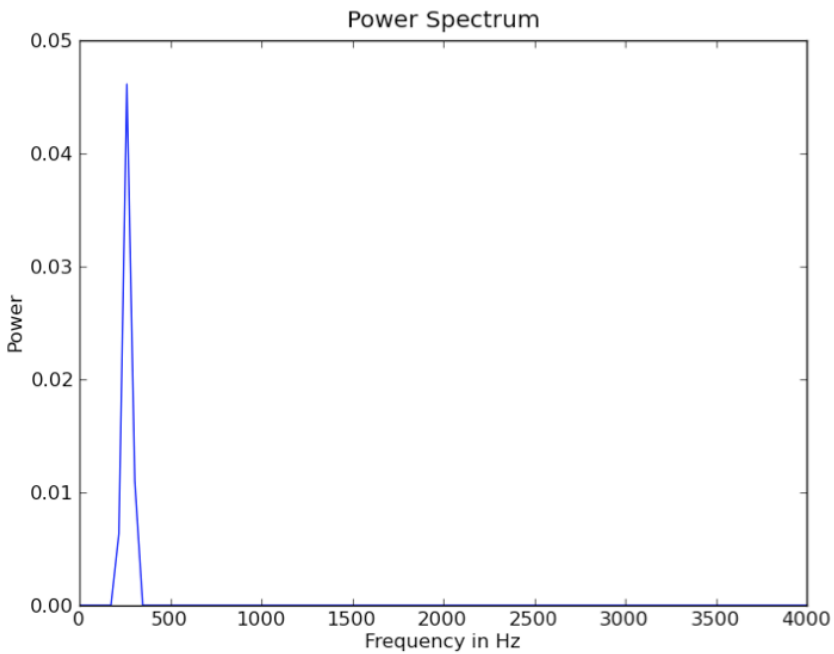


Figure: C4 Sine

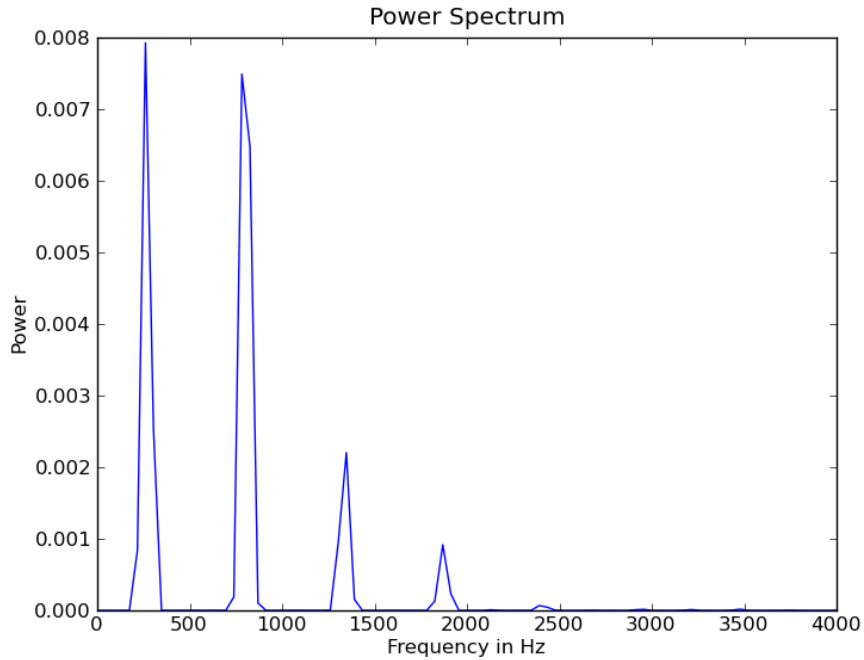


Figure: C4 Clarinet Note

Fundamental frequency (as well as pitch) will correspond to peaks in the spectrum (not necessarily the highest though).

Plotting over time

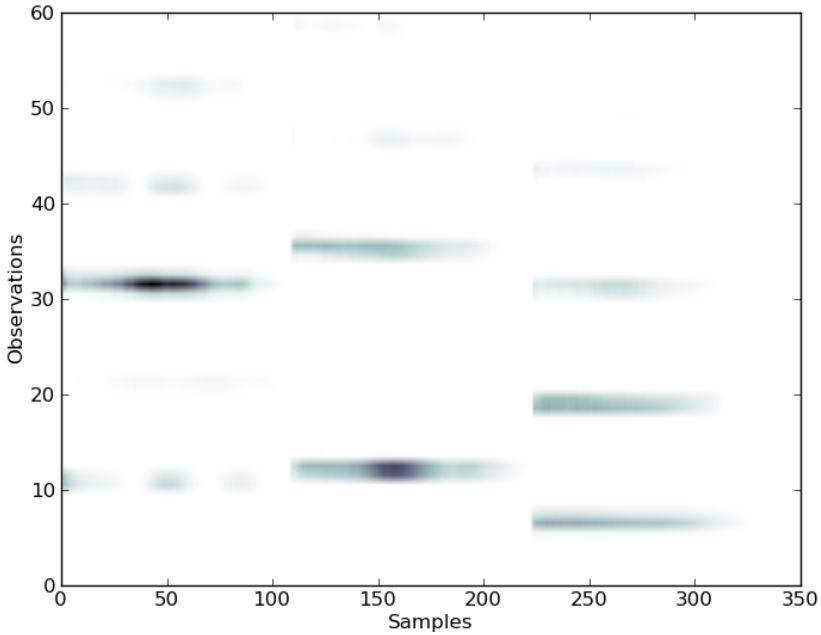


Figure: Spectrogram

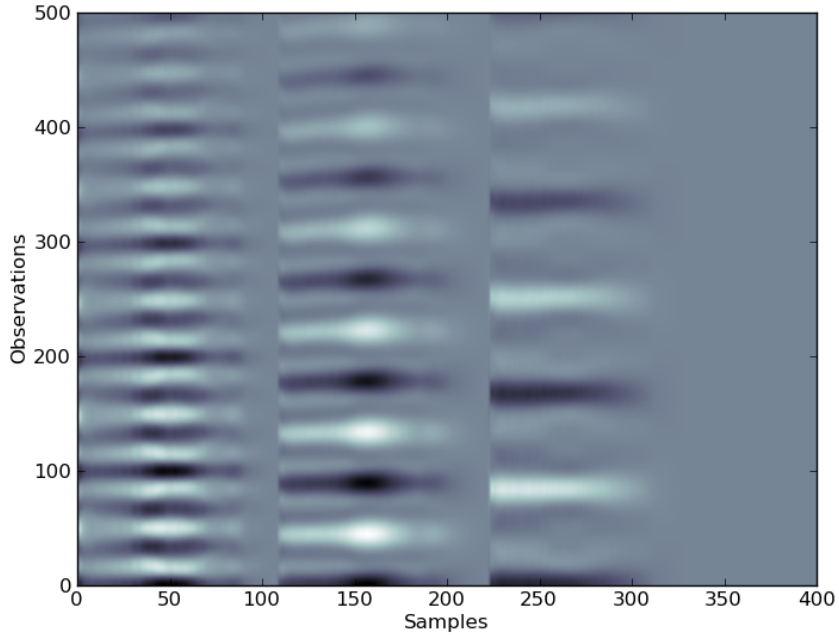
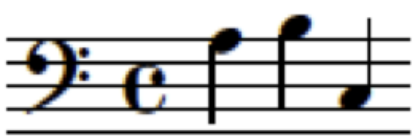


Figure: Correlogram



[Sound]

Modern pitch detection

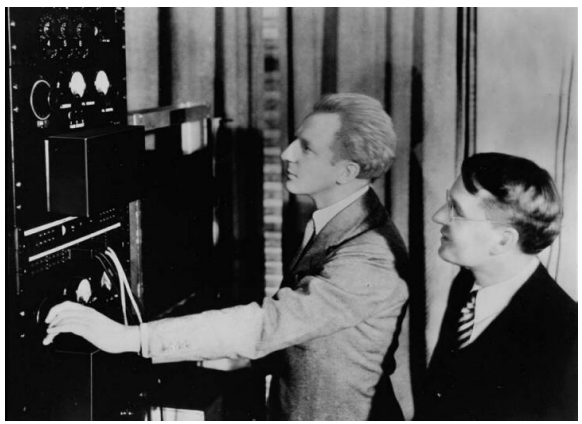
Modern pitch detection algorithm are based on the basic approaches we have presented but with various enhancements and extra steps to make them more effective for the signals of interest. Open source and free implementations available.

- YIN from the "yin" and "yang" of oriental philosophy that alludes to the interplay between autocorrelation and cancellation.
- SWIPE a sawtooh waveform inspired pitch estimator based on matching spectra

Pitch Perception

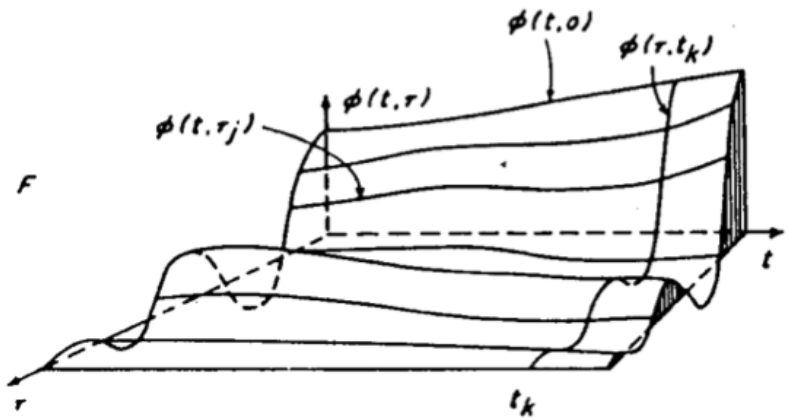
University of Victoria Computer Science

- Pitch is not just fundamental frequency
- Periodicity or harmonicity or both ?
- How can perceived pitch be measured ? A common approach is to adjust sine wave until match
- In 1924 Fletcher observed that one can still hear a pitch when playing harmonic partials missing the fundamental frequency (i.e bass notes with small radio)



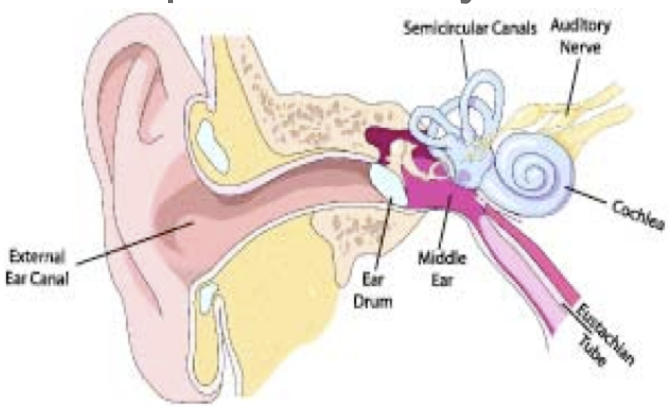
Duplex theory of pitch perception

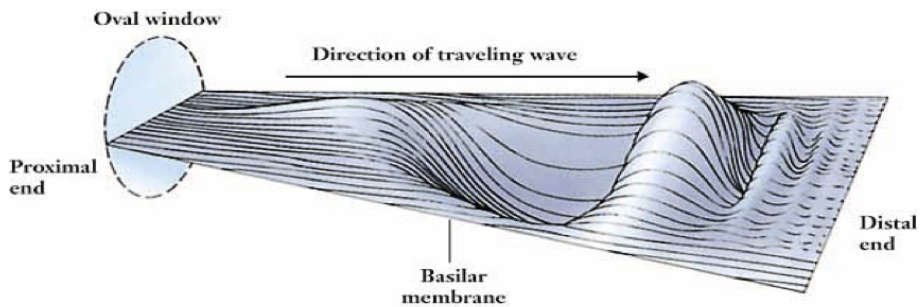
- Proposed by J.C.R Licklider in 1951 (also a realy visionary regarding the future of computers)
- One perception but two overlapping mechanisms
  - Counting cycles of a period $< 800Hz$
  - Place of excitation along basilar membrane $> 1600Hz$



The human auditory system

Incoming sound generates a wave in the fluid filled cochlea (causing the basilar membrane to be displaced - 15000 inner hair cells). Originally it was thought that the chochlea acted as a frequency analyzer similar to the Fourier transform and the perceived pitch was based on the place of highest excitation. Evidence from both perception and biophysics showed that pitch perception can not be explained solely by the place theory.





Auditory Models

From "On the importance of time: a temporal representation of sound" by Malcolm Slaney and R. F. Lyon.

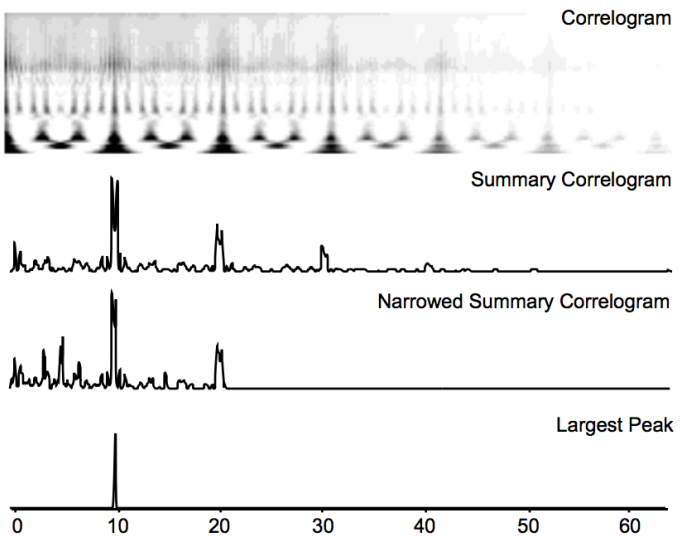
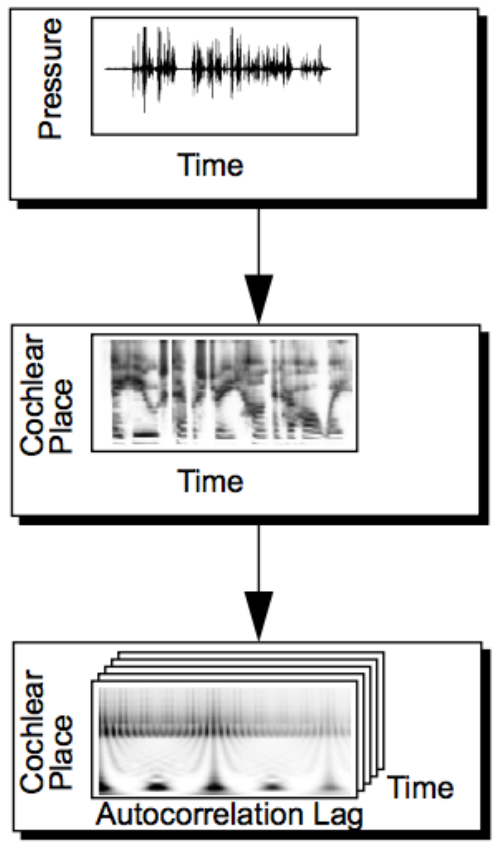


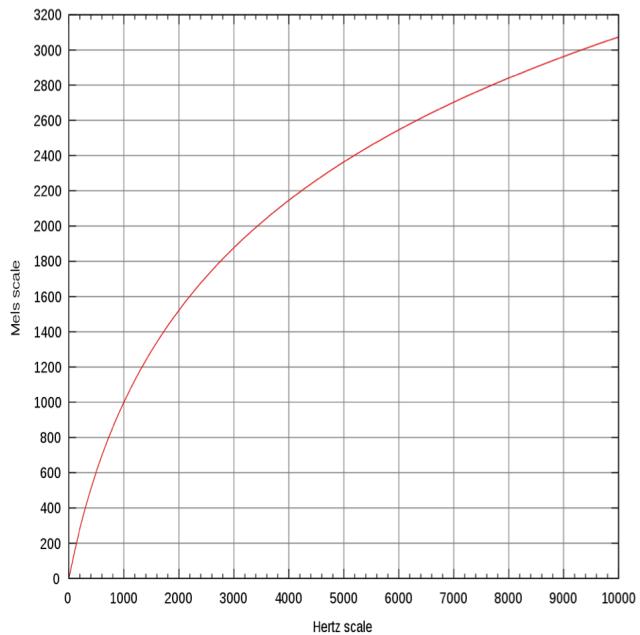
Fig. 16- Pitch of a vowel. Data processing steps in a correlogram pitch detector are illustrated he computing the summary or integrated correlogram, subharmonics are considered using narrowed autocorrelation technique. Finally, if desired, the highest peak can be chosen considered the pitch.



Perceptual Pitch Scales

University of Victoria Computer Science

- Attempt to quantify the perception of frequency
- Typically obtained through just noticeable difference (JND) experiments using sine waves
- All agree that perception is linear in frequency below a certain breakpoint and logarithmic above it, but disagree on what that breakpoint is (popular choices include 1000, 700, 625 and 228)
- Examples: Mel, Bark, ERB



Musical Pitch

University of Victoria Computer Science

- In many styles of music a set of finite and discrete frequencies are used rather than the whole frequency continuum.
- The fundamental unit that is subdivided is the octave (ratio of 2 in frequency).
- Tuning systems subdivide the octave logarithmically into distinct intervals
- Tension between harmonic ratios for consonant intervals, desire to modulate to different keys, regularlity, and presence of pure fifths (ratio of 1.5 or 3:2)

Tuning systems

- **Just Intonation** uses integer ratios that make intervals sound more consonant: $\frac{1}{1}, \frac{9}{8}, \frac{5}{4}, \frac{4}{3}, \frac{3}{2}, \frac{5}{3}, \frac{15}{8}, \frac{2}{1}$
- **Pythagorean tuning** derives all notes from perfect fifths $\frac{3}{2}$ $(\frac{1}{1}, \frac{256}{243}, \frac{9}{8}, \dots)$. Pythagorean comma (about $\frac{1}{4}$ of a semitone) reguired to get to a correct octave $\frac{2}{1}$.
- **Equal Temperament** is what is used today. All notes are spaced by logarithmically equal distances. Each "step" is higher by $\sqrt[12]{2}$ i.e to go up a step you need to multiply the current frequency by $\sqrt[12]{2} = 1.0594$

Notation

The 12 notes corresponding to each octave are mapped to white and black keys on a piano keyboard. The white keys are named using letters (A,B,C,D,E,F,G) or syllables (Do, Re, Mi, Fa, Sol, La, Ti) and the black keys are referenced using modifiers (flat # or sharp b). For example the black key to the right of a C can be referenced as either a C# or a Db.

MIDI

In order to associate each note with an actual frequency a reference tuning must be provided for one note. Today the common choice is A4 and 440Hz. MIDI (Music Instrument Digital Interface) which is a digital format for storing pitch and timing information, stores each note as an integer between 0 and 128. Converting from frequency f to MIDI note number m can be done as follows:

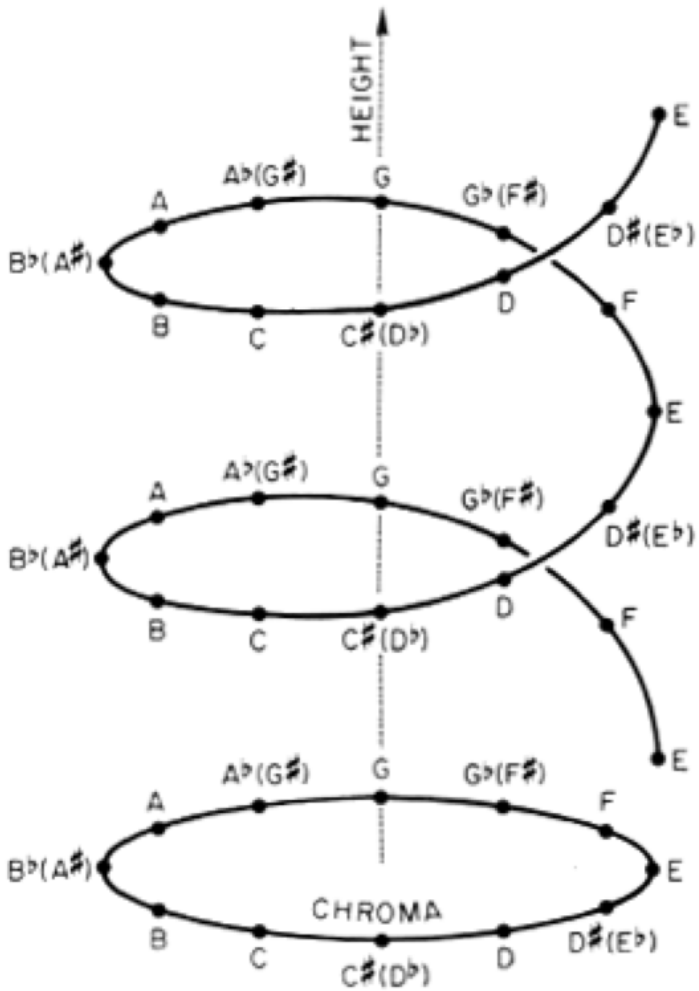
$$m = 69 + 12log_2(f/440)$$

Pitch Helix

Pitch perception has two dimesions:

- **Height**: naturally organizes pitches from low to high
- **Chroma**: represents the inherent circularity of pitch (octaves)

Linear pitch (i.e log(frequency)) can be wrapped around a cylinder to mode the octave equivalence.



From frequency to musical pitch

Sketch of a simple pitch detection algorithm

- Perform the FFT on a short segment of audio typically around 10-20 milliseoncds
- Select the bin with the highest peak
- Convert the bin index k to a frequency f in Hertz:

$$f = k * (Sr/N)$$

  where Sr is the sampling rate, and N is the FFT size.
- Map the value in Hertz to a MIDI note number

$$m = 69 + 12log_2(f/440)$$

Table of Contents I

Query by Humming (QBH)

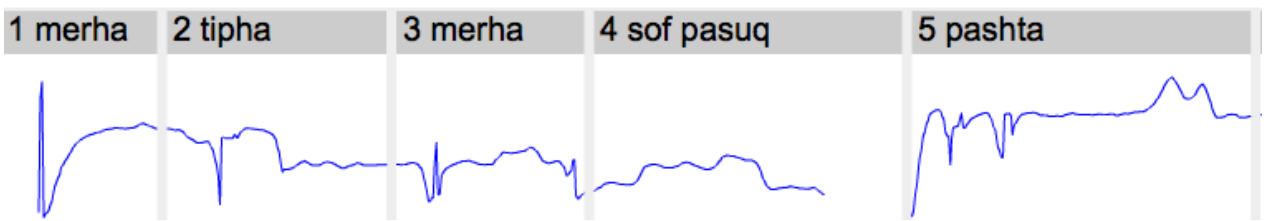
University of Victoria Computer Science

- Users sings a melody [Musart QBH examples]
- Computer searches a database of refererence tracks for a track that contains the melody
- Monophonic pitch extraction is the first step
- Many more challenges: difficult queries, variations, tempo changes, partial matches, efficient indexing
- Commercial implementation: Midomi/SoundHound
- Academic search for classical music: Musipedia

Chant analysis

- Computational Ethnomusicology
- Transition from oral to written transmission
- Study how diverse recitation traditions having their origin in primarily non-notated melodies later became codified
- Cantillion - joint work with Daniel Biro [Link]

וַיֹּאמֶר אֱלֹהִים יִקָּווּ הַמַּיִם



Summary

- There are many fundamental frequency estimation (sometimes also called pitch detection) algorithms
- It is important to distinguish between fundamental frequency, measured pitch and perceived pitch
- F0 estimation algortihms can roughly be categorized as time-domain, frequency-domain and perceptual
- Query-by-humming requires a monophonic pitch extraction step
- Chant analysis is another more academic application