LING 490 - SPECIAL TOPICS IN LINGUISTICS

Fundamentals of Digital Signal Processing

Yan Tang

Department of Linguistics, UIUC

Week 14

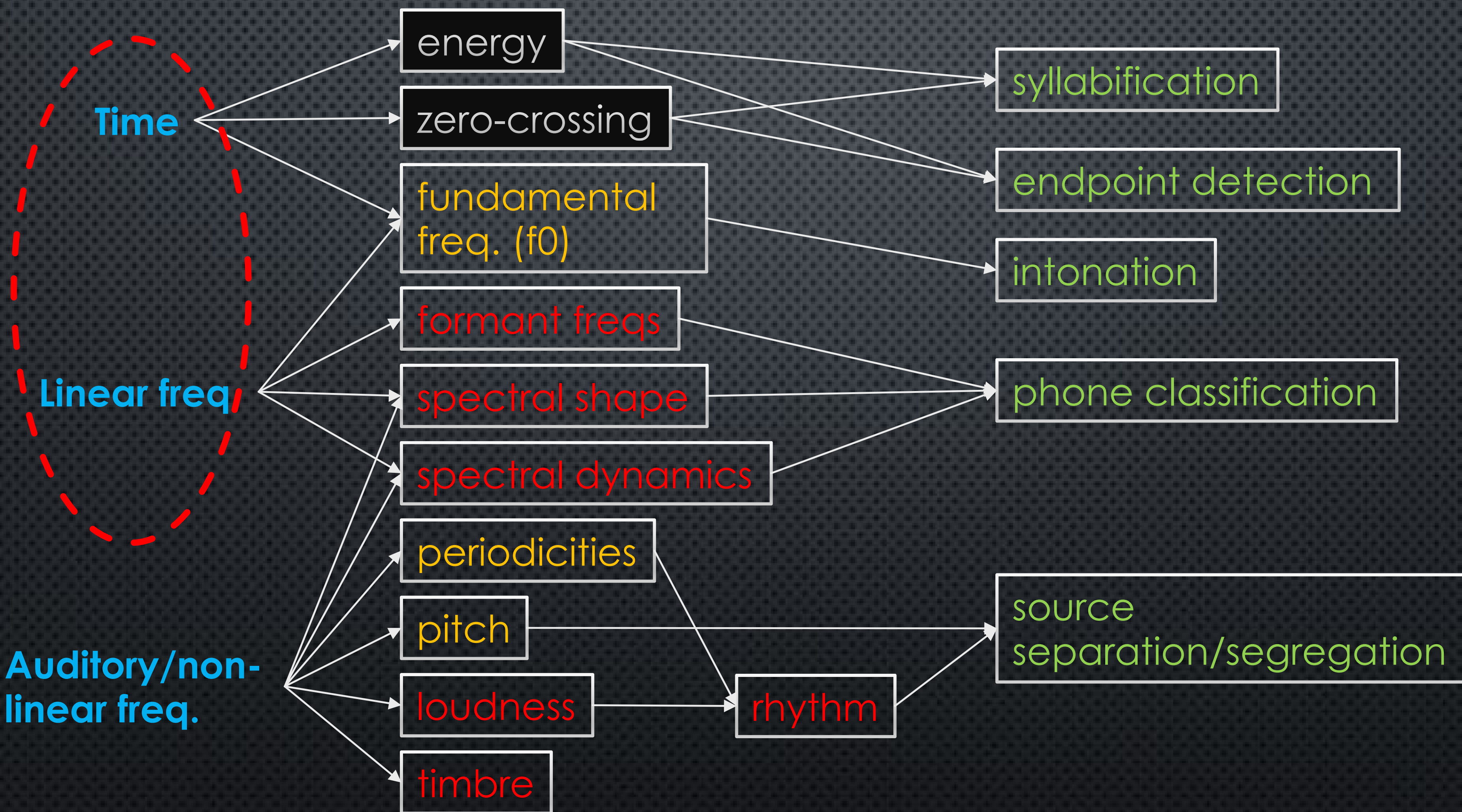
Last week…

- The domains in which speech parameters are extracted
  - Time, frequency and auditory
- Quantifying the "size" or magnitude of a signal
  - Energy
  - Magnitude function
- Quantifying the the strength of a signal in a unit time
  - Power
  - Intensity

Last week…

- Zero-crossing rate: the rate at which the signal changes from positive to negative
- For speech signals:
  - Voiced regions: ***high*** energy, ***low*** zero-crossing rate
  - Unvoiced regions: ***low*** energy, ***high*** zero-crossing rate
- Applications: endpoint detection, voiced-unvoiced detection

Useful parameters of speech signals



Pitch - definition

- Definition: ***Pitch** is that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale*. (American Standards Association, 1960)

- Definition: ***Pitch** is that auditory attribute of sound according to which sounds can be ordered on a scale from low to high*. (American National Standards Institute, 1994)

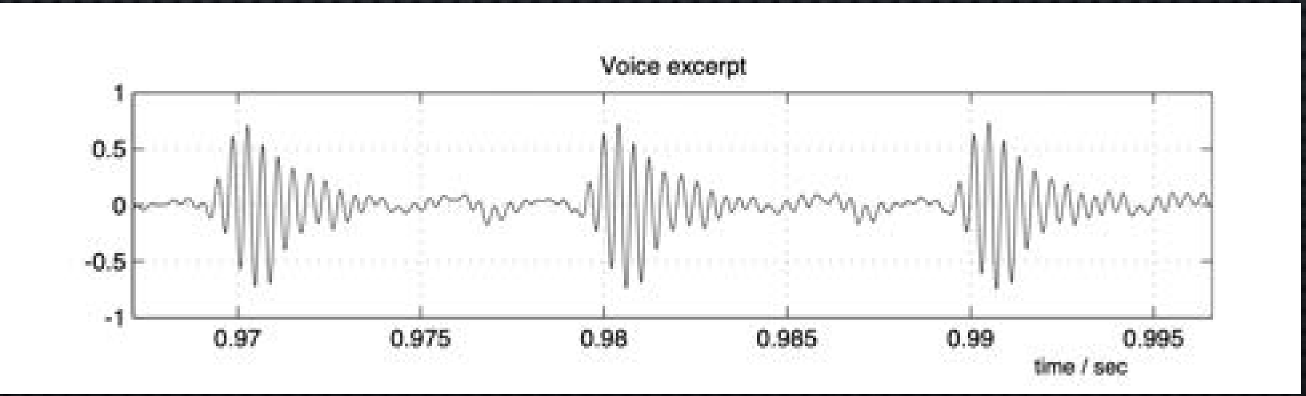
Pitch - definition

"…a sensation is the simplest form of cognition. It is a simple impression produced in the mind by a stimulus…" (psychologydiscussion.net)

- Pitch is an attribute of sensation
- The word *pitch* should not be used to refer to a physical attribute of a sound

Pitch - introduction

- Pitch is related to the repetition rate of a signal
- Repetition rate of a sinusoid is its frequency
- Repetition rate of a complex tone (i.e. with harmonics) is its fundamental frequency (F0)



Why is pitch important?

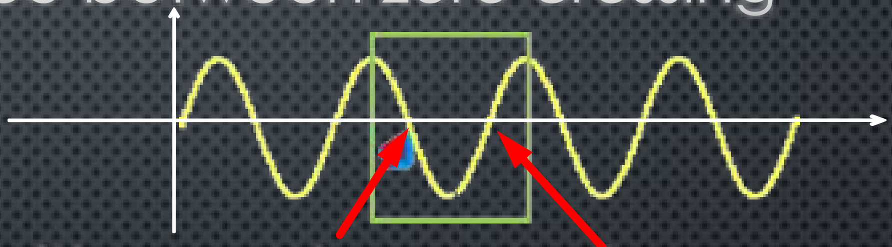
- Conveys meaning in language
- Cue to determining number of sources in a mixture and then separating them
- Music
- Cue to speaker gender (and age)
- Speaker id and verification

Pitch estimation

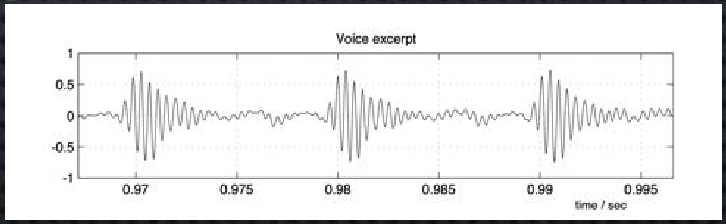
- Pitch estimation is really **F0** estimation
- Assumes the signal has some periodic content
- Basic goal is to identify how often the signal (or portion of) repeats itself
- **Remember:** signal may contain many different periodicities

A basic approach

- Recall ZCR - simple approach would be to measure the distance between zero crossing points of the signal.



- Does not work well with complex waveforms which are composed of multiple sine waves with differing periods (and noise).



- Nevertheless, the algorithm's simplicity makes it 'cheap' to implement if the target signal suits.

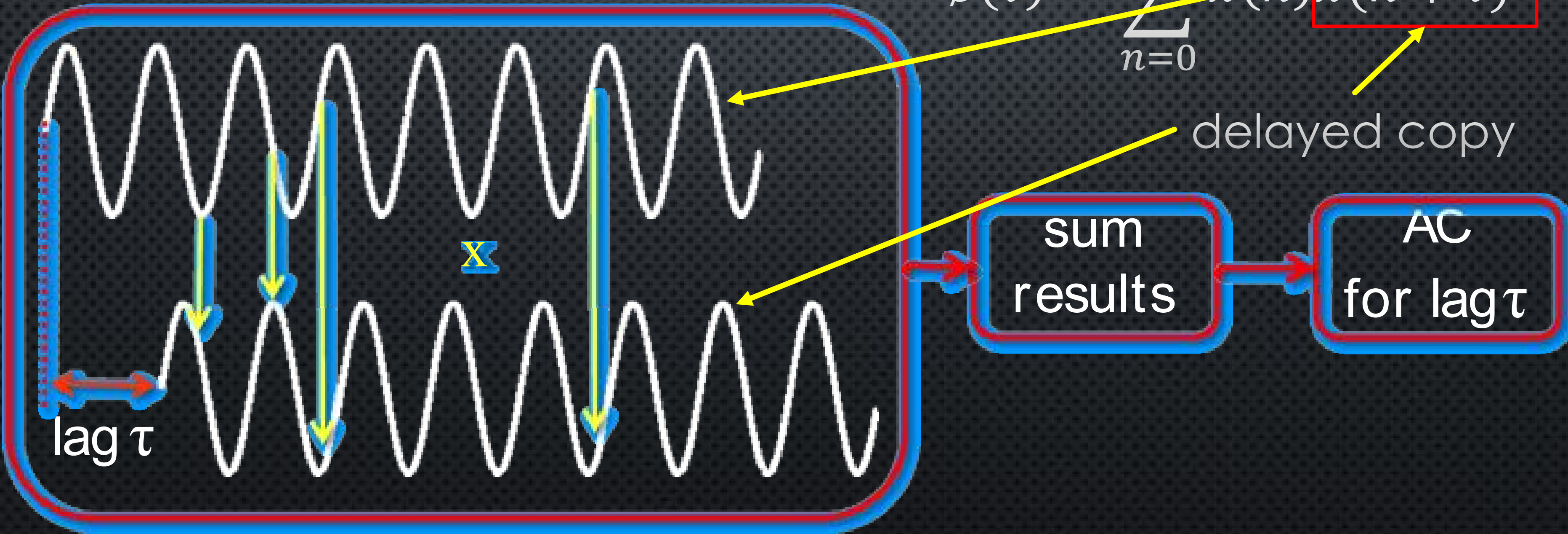
Autocorrelation

- The "classic" pitch estimation approach
- Correlation is the mutual relationship between two or more random variables
- Autocorrelation is the correlation of a signal with itself
- Many different definitions of autocorrelation are in use and not all of them are equivalent

Autocorrelation

- Common aspect is the multiplication of $x(n)$ with a time-delayed version of itself then sum values.

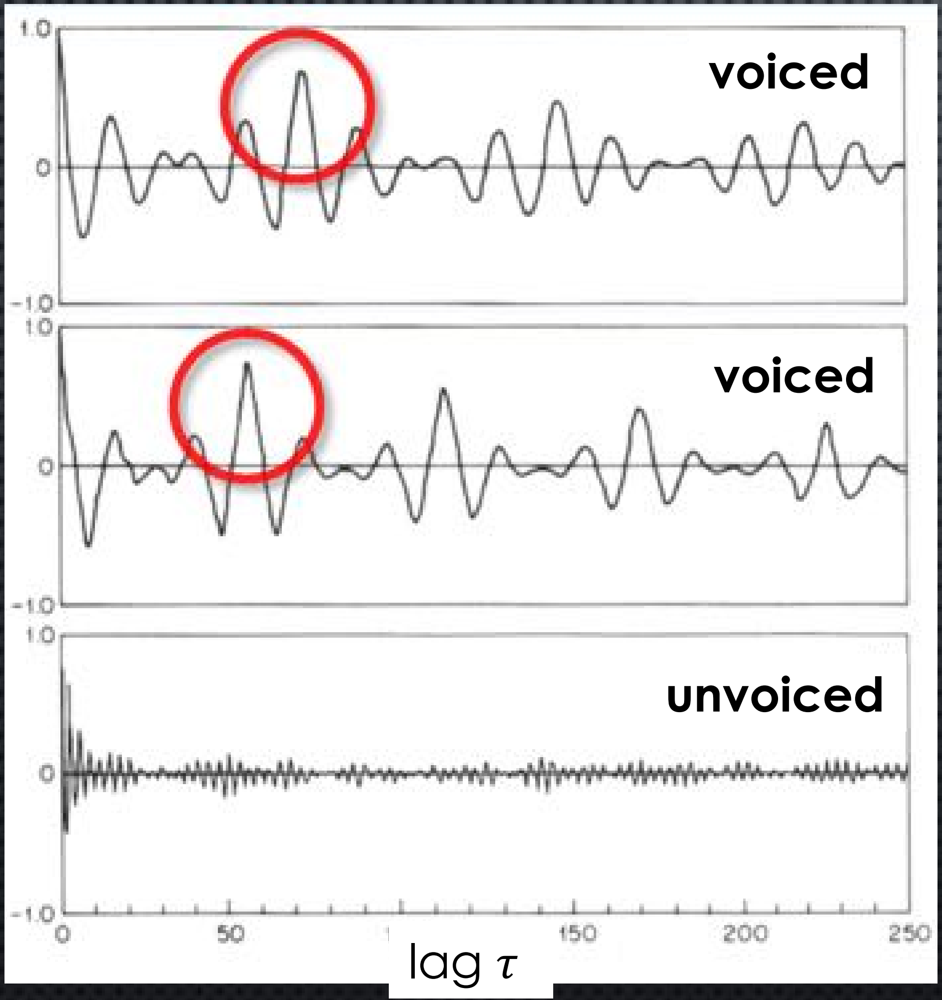
$$\emptyset(\tau) = \sum_{n=0}^{N-1} x(n)\boxed{x(n+\tau)}$$



Autocorrelation

- Performs this at a range of different time-delays
- Produces another signal
- Peaks relate to periodicities in $x(n)$

Autocorrelation



Autocorrelation using a rectangular window with N = 400

- Note shrinking peaks due to wave shape constantly changing.
- Periods further away are increasingly different therefore show lower correlation.

Autocorrelation and windows

- Equation on early slide conveniently left out the windowing term.

$$\emptyset(\tau) = \sum_{n=0}^{N-1} x(n)x(n+\tau)$$

- Autocorrelation is performed using two windowed sections of $x(n)$

- Different *window functions* can make a difference

- Different *window sizes* can also make a difference

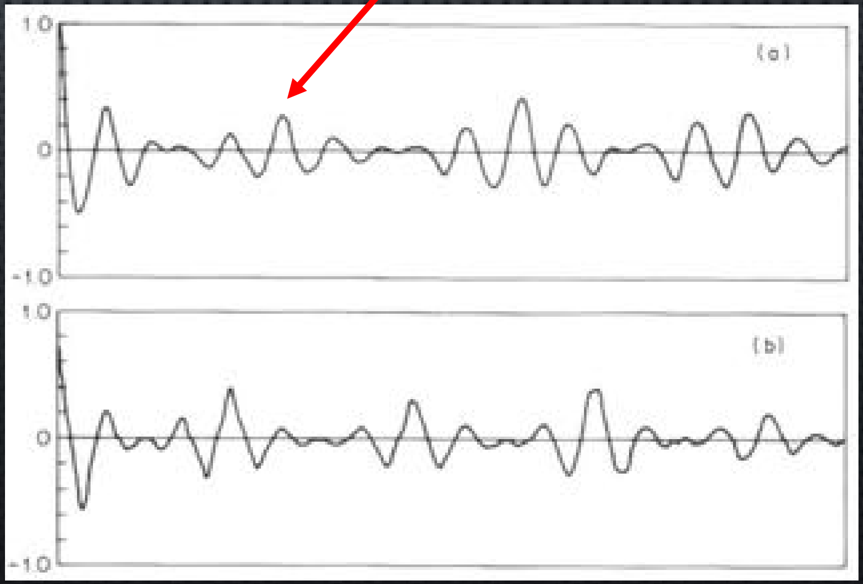
Window type

- Different windowing functions produce different autocorrelation outputs

Hamming window introduces 'tapering' of the speech signal, therefore weaker indication of periodicity

Rectangular

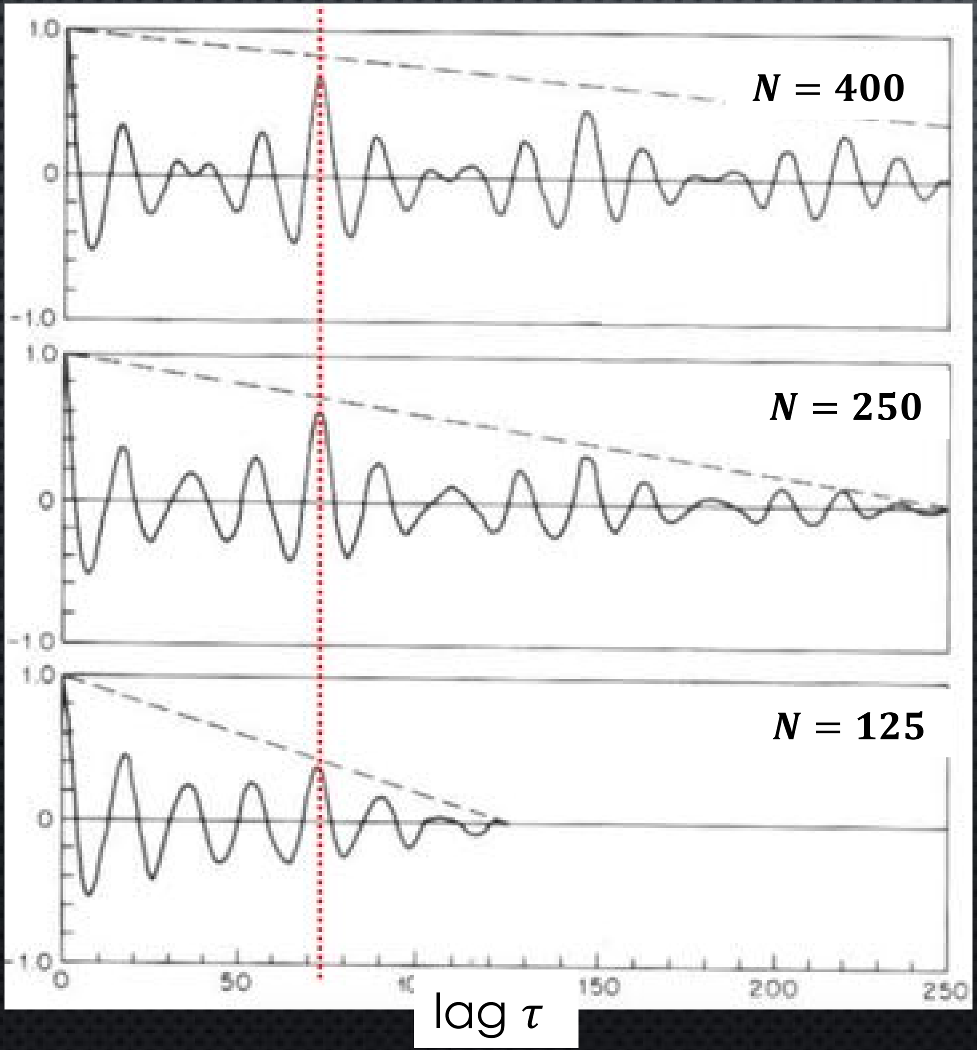
Hamming



What’s the best window size?

- Because of the changing properties of the speech signal, N should be as small as possible.
- However, N must be large enough to cover at least two periods.

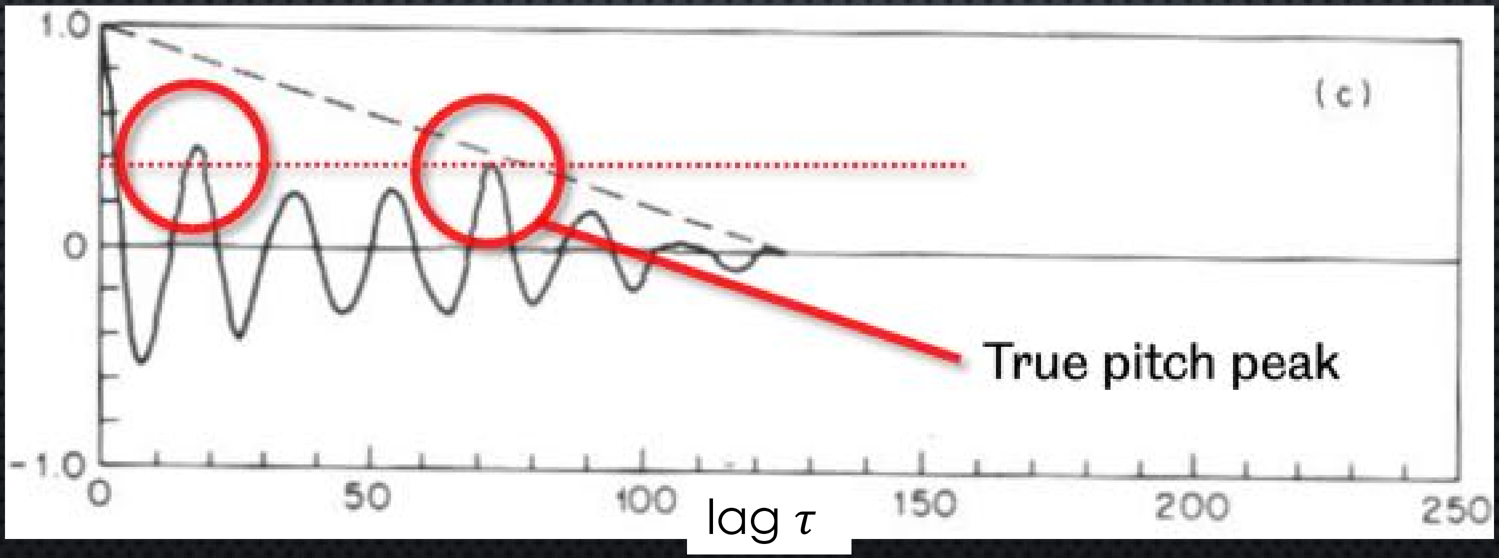
Window size comparison



- Dotted line shows upper bound on peak height.
- Strong height difference between pitch peak and others in top 2 panels.
- Window contains less than 2 fundamental periods – weaker peak.

Issues of autocorrelation

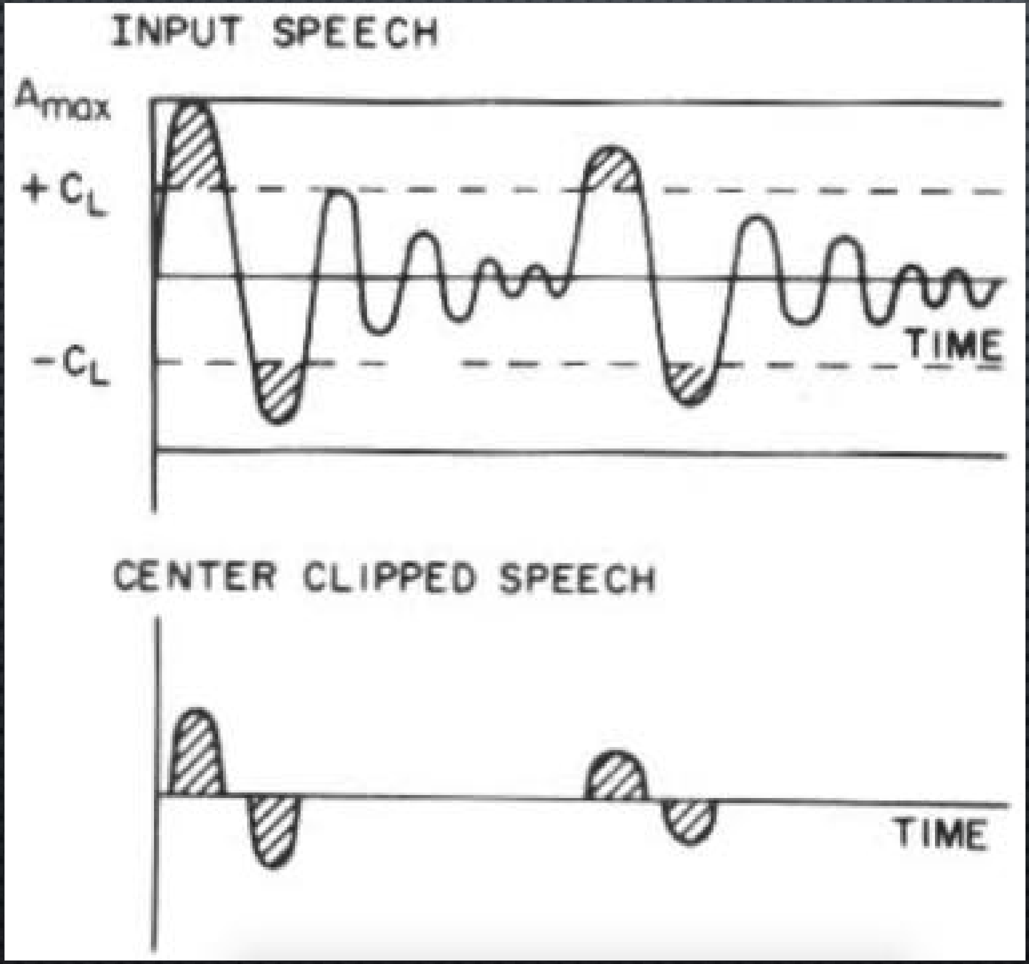
- Autocorrelation can have too many peaks.
- Retaining too much info from original signal.
- If window short of pitch period, such peaks can be bigger than pitch period peak!



Preprocessing before detection

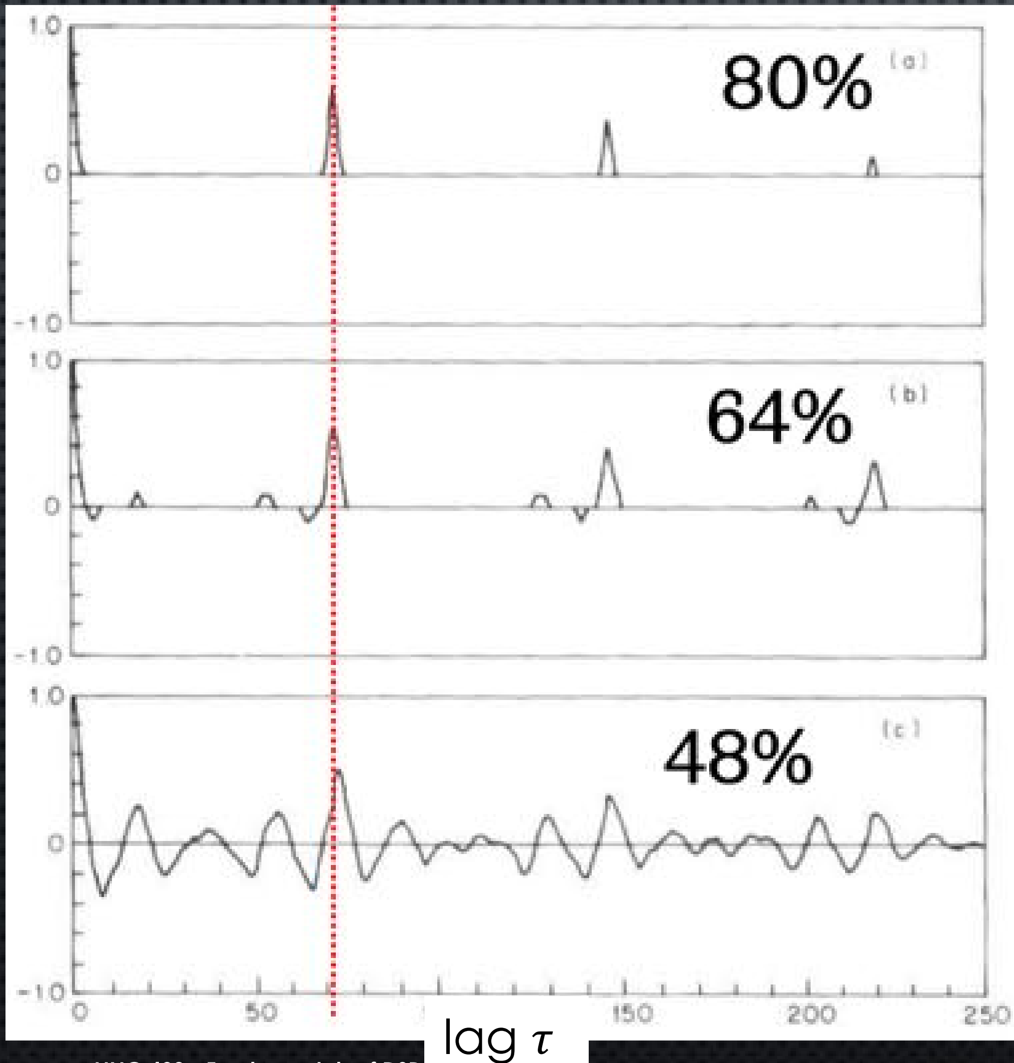
- Try and make periodicity more prominent
- ‘Spectrum flattener’ – remove effects of vocal tract
- Classic approach is the nonlinear transformation called ‘***centre clipping***’

Centre clipping



- Only retain parts of signal which are above a percentage C_L of the maximum amplitude A_{max}

The effect of centre clipping



- High clipping level provides clearest indication of periodicity

- However, care must be taken with signals whose amplitude varies – low amplitude sections may be completely removed!

Common issues

+ Good accuracy for highly periodic signals

- False detection problems (often ‘octave errors’)

- Cope badly with noisy signals (depending on the implementation)

- Basic implementations do not deal well with polyphonic sounds (which involve multiple pitches)

State-of-the-art

- Current time-domain pitch detector algorithms tend to build upon the basic methods
- Additional refinements bring the performance more in line with a human assessment of pitch
  - The YIN algorithm (de Cheveigné, 2002)
- Other methods:
  - TANDEM-STRAIGHT (Kawahara et al, 2011)
  - Glottal Inverse Filtering (Raitio et al, 2011)
  - …