LING 490 - SPECIAL TOPICS IN LINGUISTICS

Fundamentals of Digital Signal Processing

Yan Tang

Department of Linguistics, UIUC

Week 13

Last week…

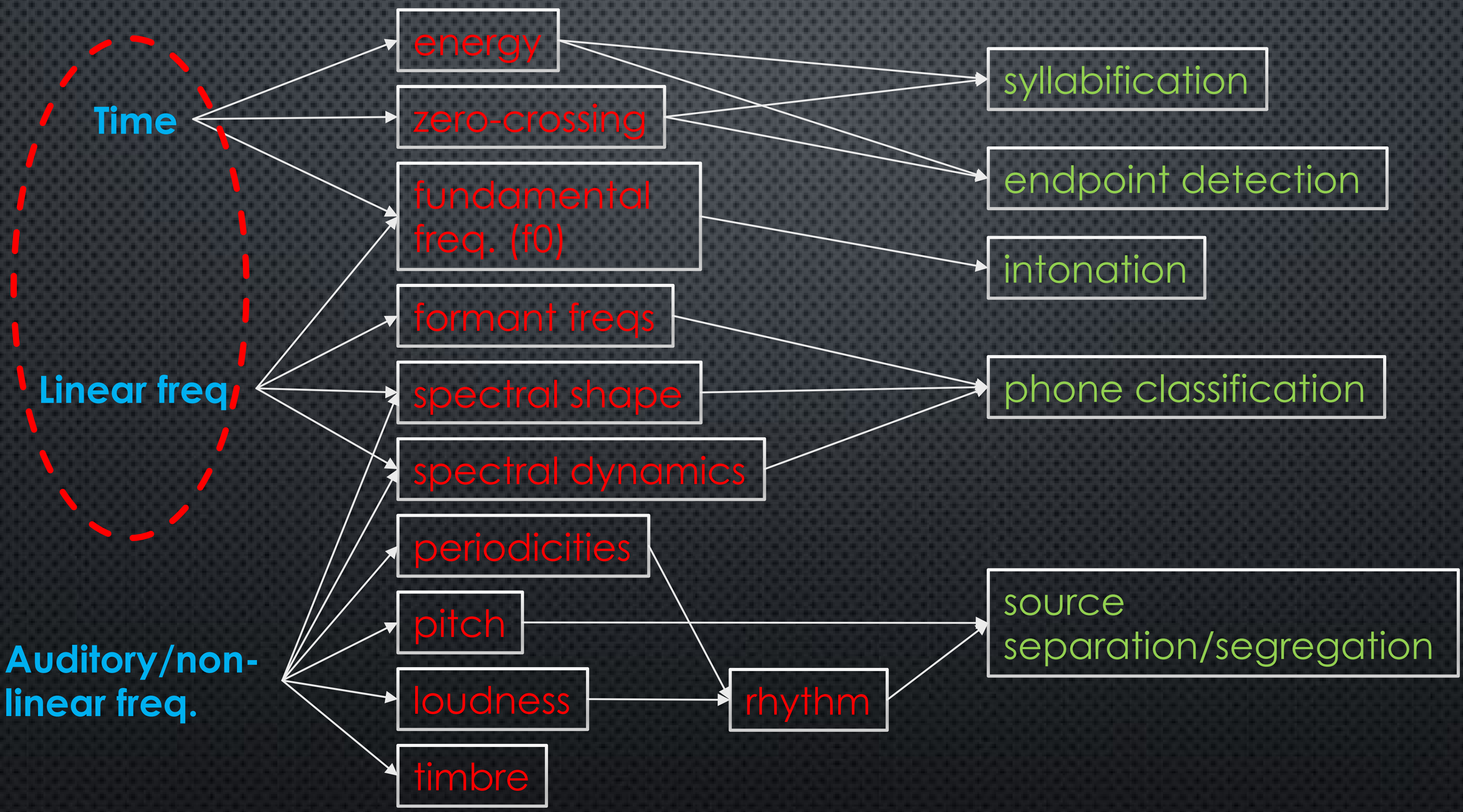
- A system formed by any number of LTI systems in ***parallel*** is an LTI system
  - The frequency domain: $R(f) = R_1\ (f) + R_2\ (f) + \cdots + R_n\ (f)$
  - The time domain: $H(t) = H_1\ (t) + H_2\ (t) + \cdots + H_n\ (t)$

- A system formed by any number of LTI systems in cascade is also an LTI system
  - The frequency domain: $R(f) = R_1\ (f) \cdot R_2\ \ (f) \cdot \cdots \cdot R_n\ \ (f)$
  - The time domain: $H(t) = H_1\ (t) \otimes H_2\ (t) \otimes \ldots \otimes H_n\ (t)$

Last week…

- The vocal tract can be modelled as a chain LTI system
  - Difference sounds due to varying shape
- Formants: characteristics of frequency response of the vocal tract
  - Formant frequency and half-power bandwidth
- Phase response of LTI systems
  - $D(f) = P_O(f) - P_I(f)$
  - An ideal phase response: linear phase response
- The *amplitude spectrum of an impulse response* of a system is the *frequency response* of the system

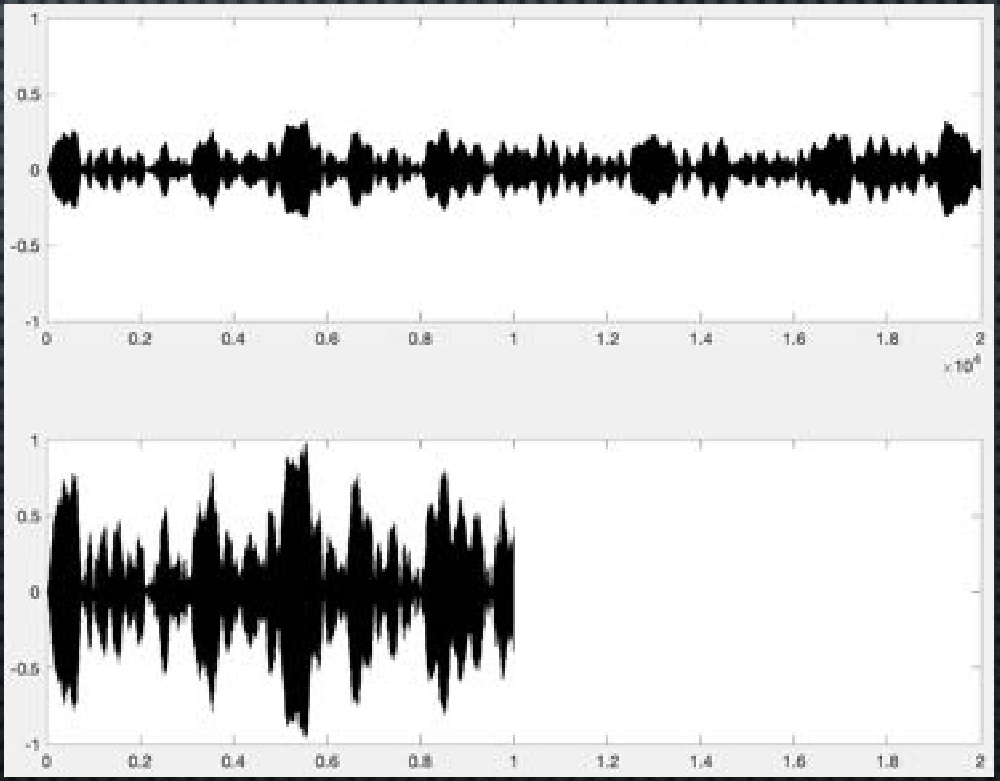
Useful parameters of speech signals



Energy

- The idea of the "size" of a signal is crucial to many applications
  - It is nice to know how much electricity can be used in a defibrillator without ill effects.
  - It is also nice to know if the signal driving a set of headphones is enough to create a sound

which signal is “bigger”?



Energy

- While both examples are digital signals, they are clearly very different signals with very different properties, e.g. duration and amplitude

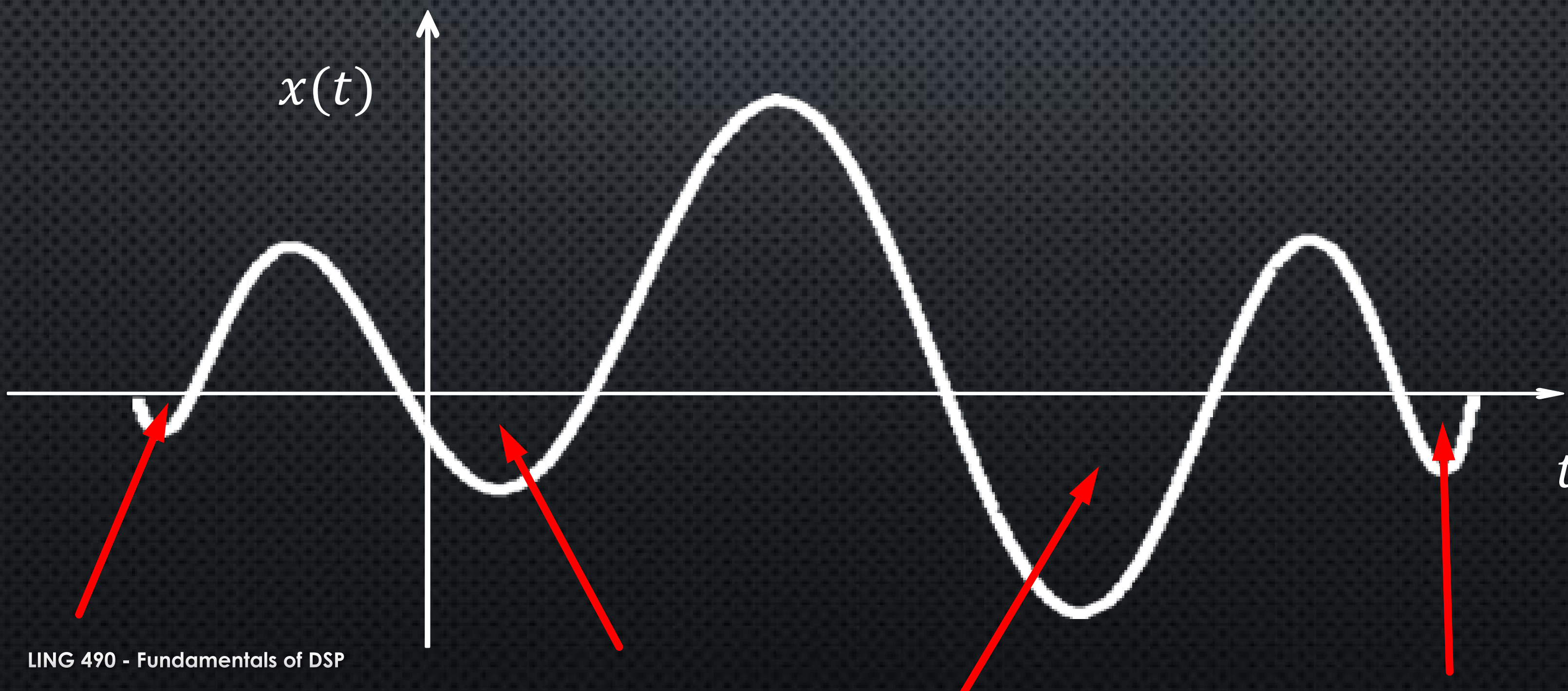
- For this reason, it is convenient to quantify this idea of "size"
  - Signal energy
  - Signal power

Energy

- We often think of signal as a function of varying amplitude through time

Energy

- Area under the curve?
- This area may have a negative part
- This negative part does not have less strength than a positive signal of the same size



Energy

- Therefore, square the signal, then find the area under that curve

Animation

Definition: Energy = area under squared signal

Energy: calculation

- For a discrete time domain signal x, energy E is defined as:

$$E = \sum_{n=0}^{N-1} x^2(n)$$

N: the number of samples

Energy: sound energy in air

- The energy of a sound in air is defined as:

$$E^s = \frac{\Delta t}{\rho c} \sum_{n=0}^{N-1} x^2(n), \qquad (Joule/m^2)$$

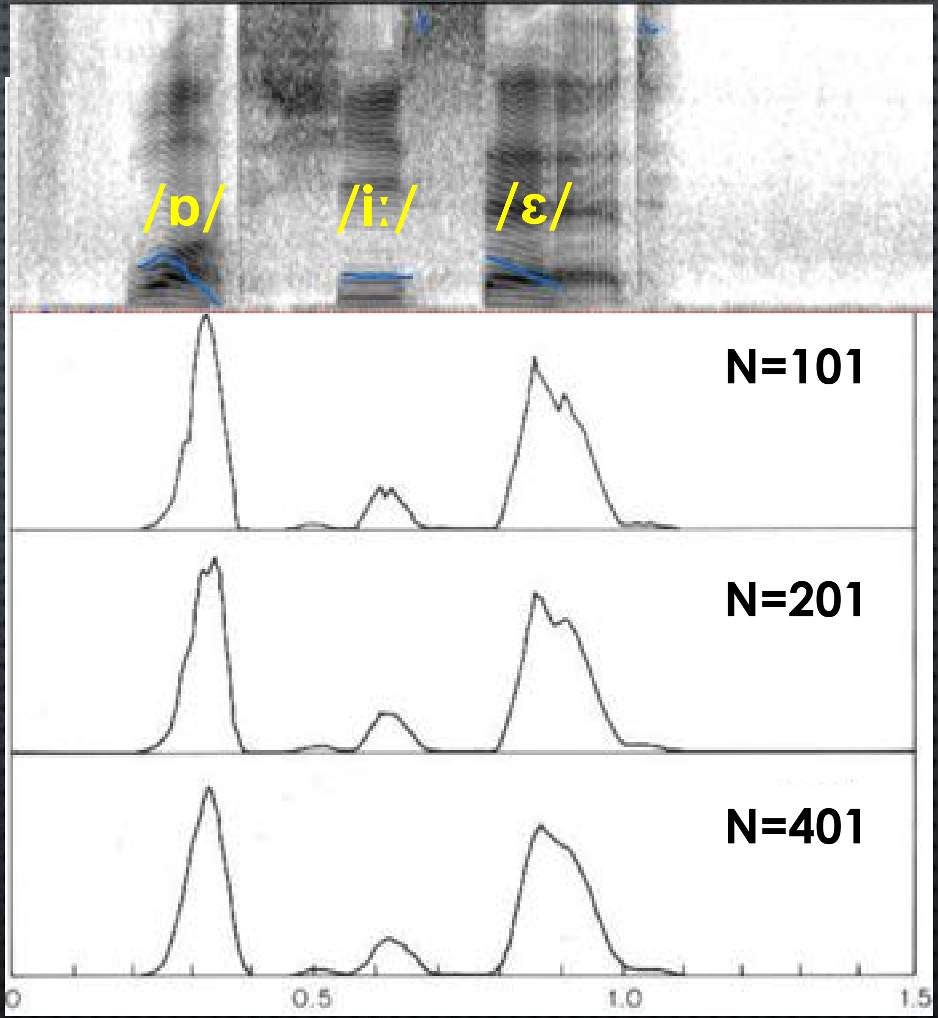
ρc: the acoustic impendence

Δt: the sampling period, i.e. $(1/sampling\ frequency)$

Energy

- Higher energy at voiced regions
- Note how measure changes as window size increases - smeared
  - larger window size, less temporal details

"what she said" in rectangle windows



time (s)

Short-time energy for rectangle windows of various length

Magnitude function

- Energy is very sensitive to large signals levels (why?)
- Can (over) emphasise large sample-to-sample variations in $x(n)$
- Magnitude function (MF) is the same as energy except the absolute value instead of the square is taken

Magnitude function: calculation

- For a discrete time domain signal x, average magnitude M is defined as:

$$M = \sum_{n=0}^{N-1} |x(n)|$$

N: the total number of samples

Power

- We have seen how to derive the energy of a signal
- What happens if we wish to compare the energies of two different signals?
  - A fair comparison if they have the *same length*
- How to compare signals with different lengths?

Mean power: calculation

- Power is a time average of energy (energy per unit time), i.e. normalised by length :

$$P = \frac{1}{N}\sum_{n=0}^{N-1} x^2(n)$$

N: the number of samples

Recall: $I \propto P$, where I is intensity

Intensity: sound energy in air

- The intensity of a sound in air is defined as:

$$I^s = \frac{1}{\rho c \cdot N} \sum_{n=0}^{N-1} x^2(n), \qquad (Watt/m^2)$$

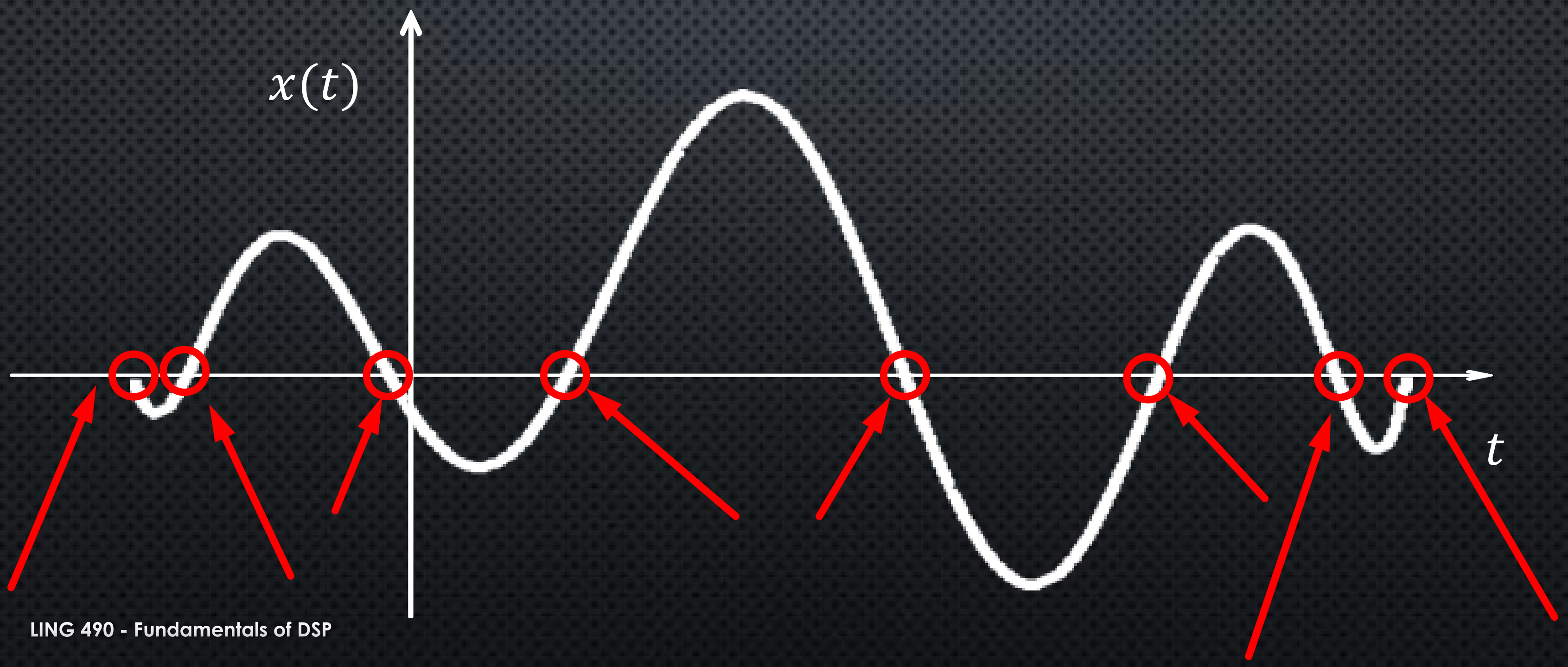
ρc: the acoustic impendence

- Sound Intensity Level (SIL) (as in Week 5)

$$SIL = 10 \cdot log_{10} \frac{I^s}{10^{-12}}$$

Zero crossing rate (ZCR)

- ZCR is the rate of sign-changes along a signal
  - i.e., the rate at which the signal changes from positive to negative or back



ZCR

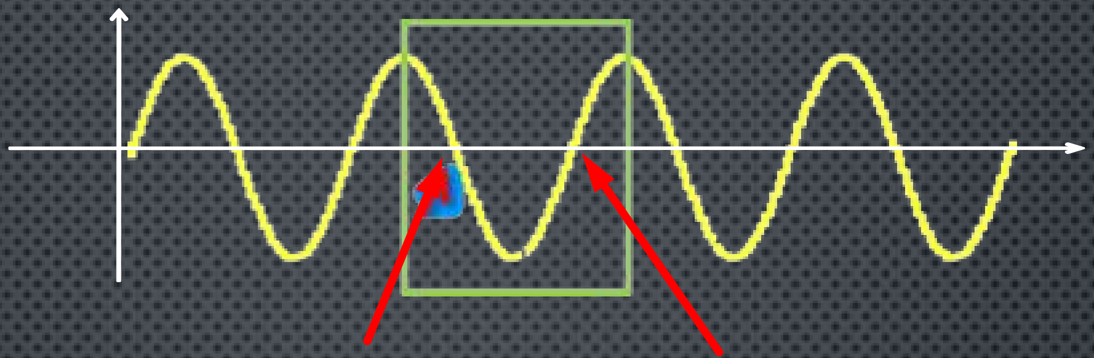
- ZCR is defined as the ratio between the number of zero-crossing times and the number of data points

$$zcr = \frac{1}{2N} \sum_{n}^{N-1} \left| \mathrm{sign}\big(x(n)\big) - \mathrm{sign}\big(x(n-1)\big) \right|$$

$$\mathrm{sign}(x(n)) = \begin{cases} 1 & if\ x(n) \geq 0 \\ -1 & if\ x(n) < 0 \end{cases}$$

N: the number of samples

ZCR

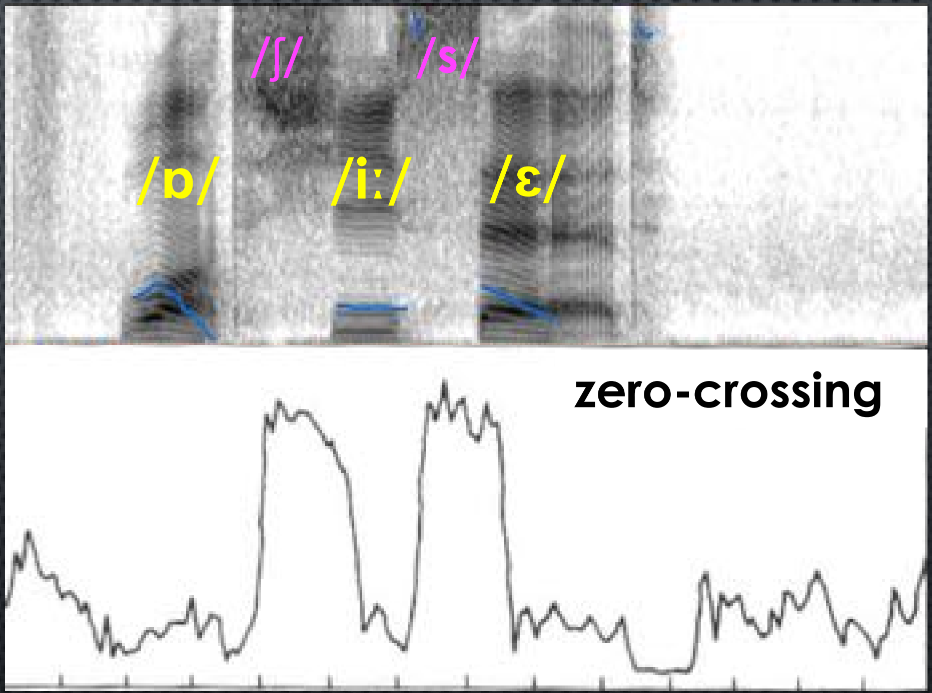


- Can be considered a measure of frequency content – especially for narrowband signals
  - e.g. sinusoid has 2 zero crossings in each period
- However, speech is broadband; interpretation of ZCR becomes less clear.
  - Low ZCR: low frequency content – voiced?
  - High ZCR: high frequency content – unvoiced?
  - What's boundary between low and high? Good question!

ZCR

- High ZCR
  - Higher frequencies
  - Unvoiced speech
- Low ZCR
  - lower frequencies
  - Voiced speech

"what she said"

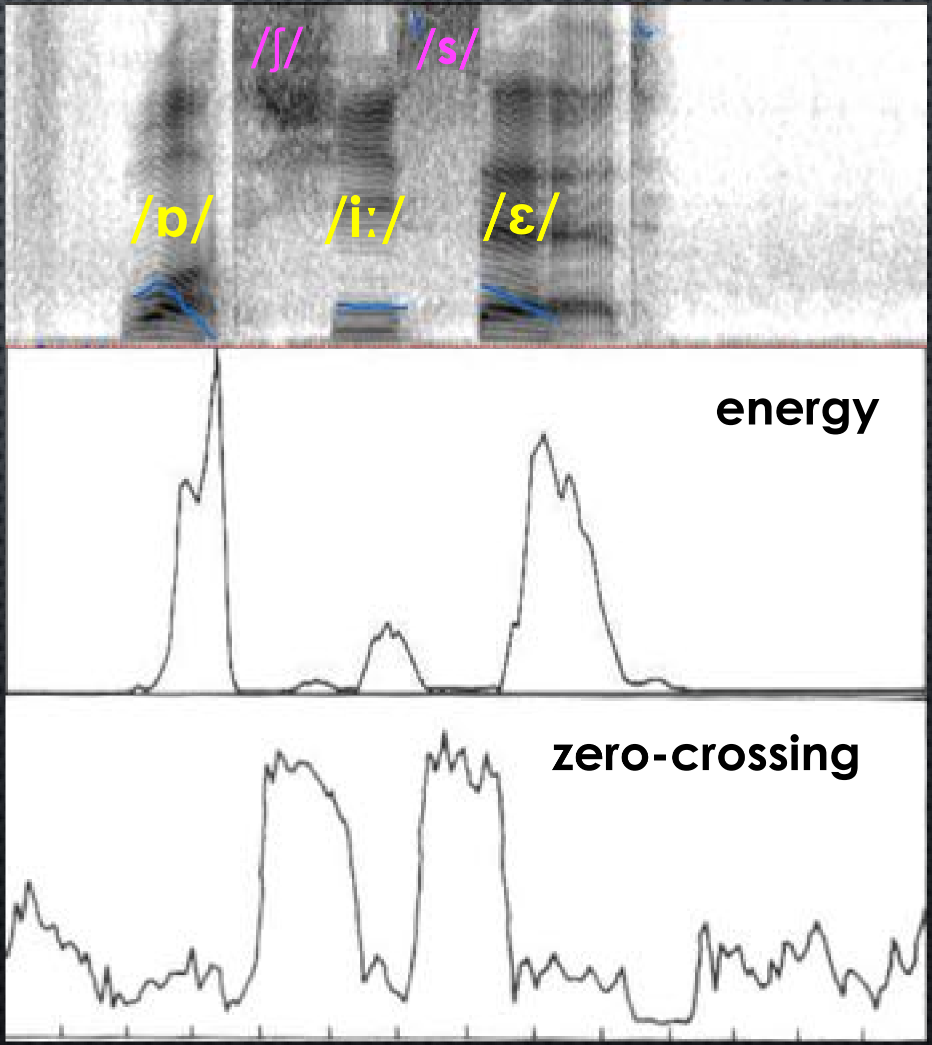


time (s)

Energy and ZCR

- Interesting property
  - In voiced regions, energy is high and ZCR is low (in general!)
  - In unvoiced regions: opposite patterns
- Implication
  - Endpoint detection
  - Voiced/unvoiced detection

"what she said"

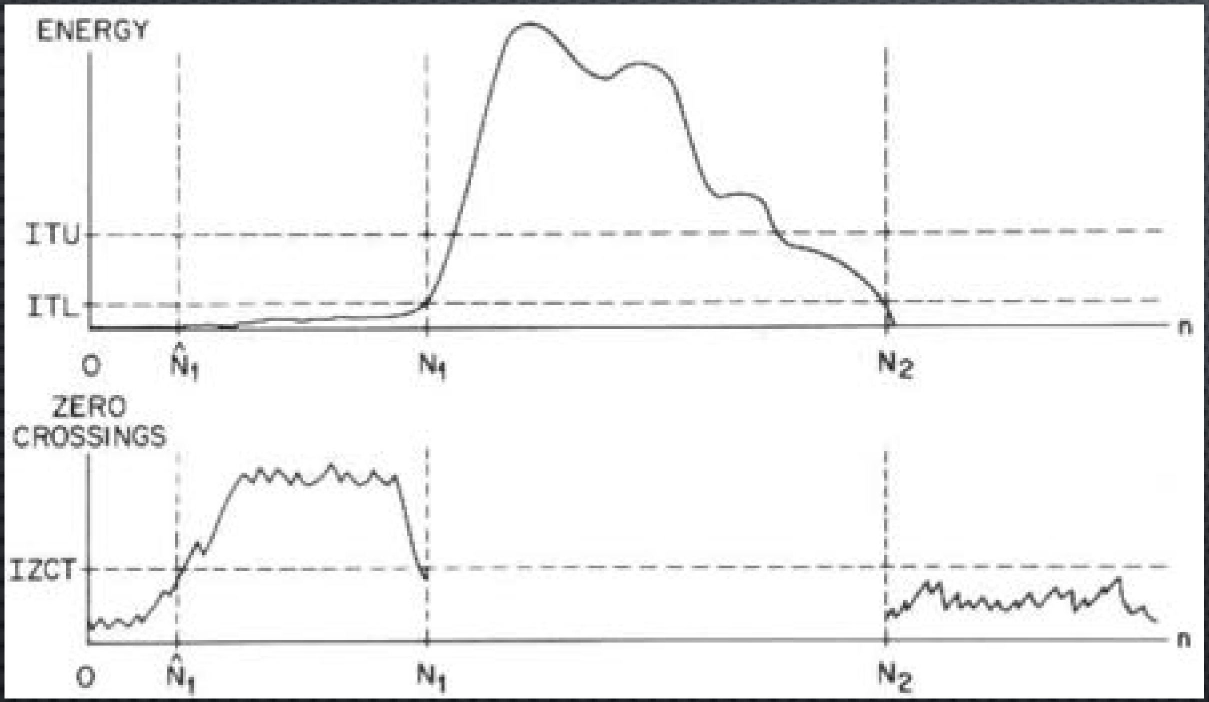


time (s)

Endpoint detection

- Often useful to be able to determine beginning and end of an utterance
- Limits unnecessary processing of silence; makes processing more consistent
- A simple approach: energy and ZCR

Endpoint detection



Principles:

- Rough estimates of utterance boundaries are where energy goes above certain threshold
- If ZCR is above another threshold, extend the estimated boundary to where ZCR falls below the threshold