Lecture 3: Fundamentals: Signal processing, acoustics, speech signals

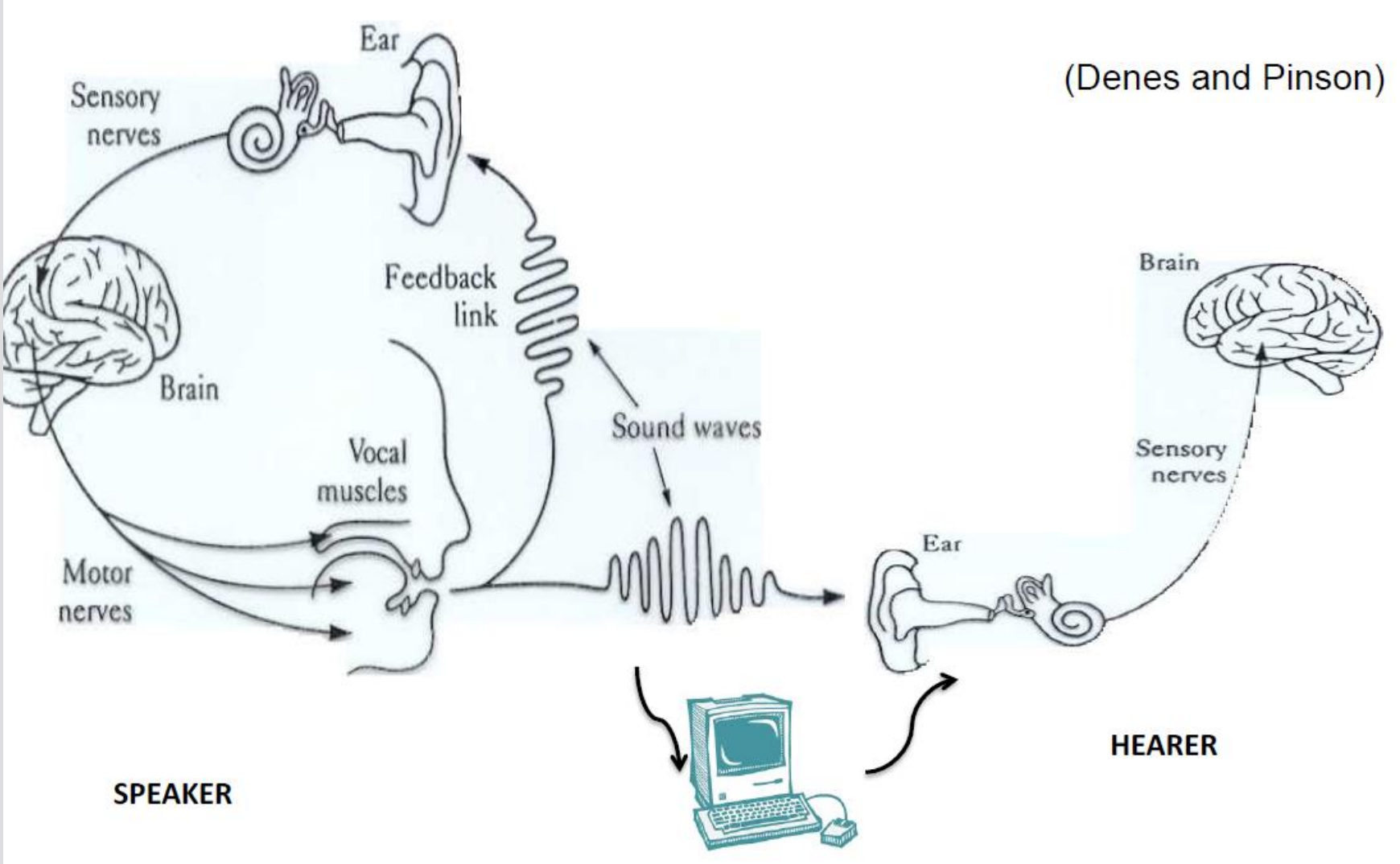
http://www.chatterboxspeech.com.au/speech-articulation/

Waldo Nogueira

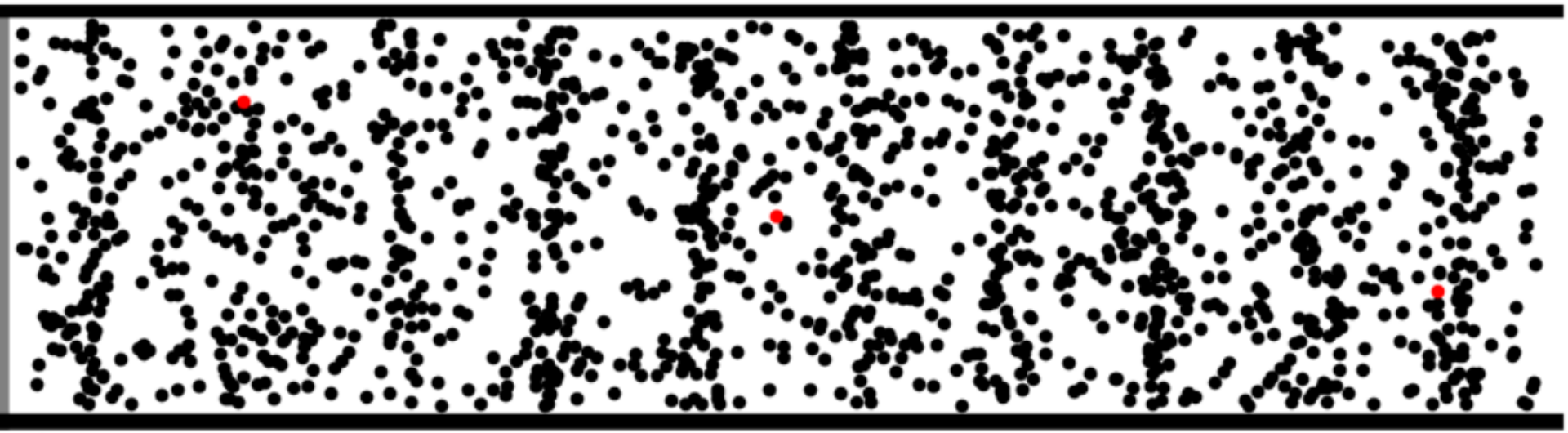
04/10/2022

Medizinische Hochschule Hannover
DHZ Deutsches HörZentrum Hannover
VIANNA
Hearing 4all

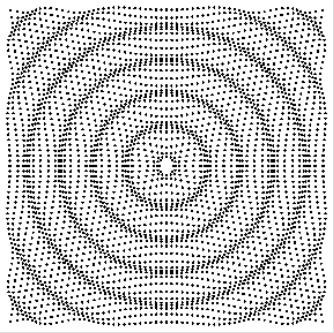
Speech Chain



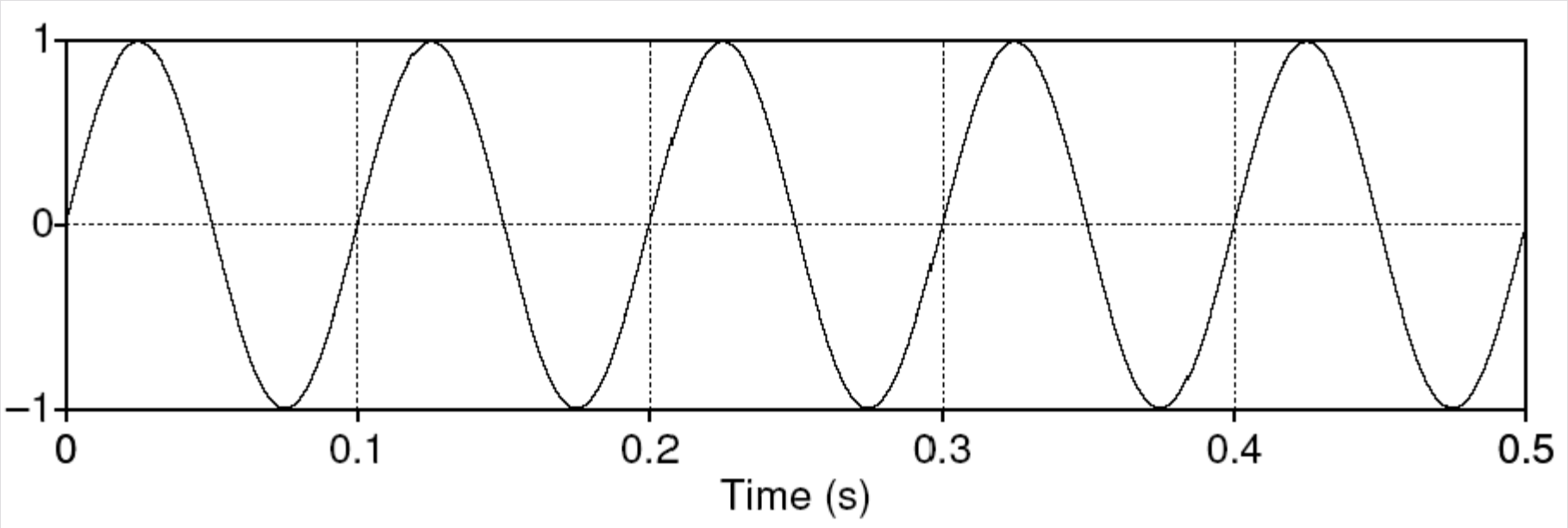
Sound Wave



- Animation courtesy of Dr. Dan Russell, Kettering University

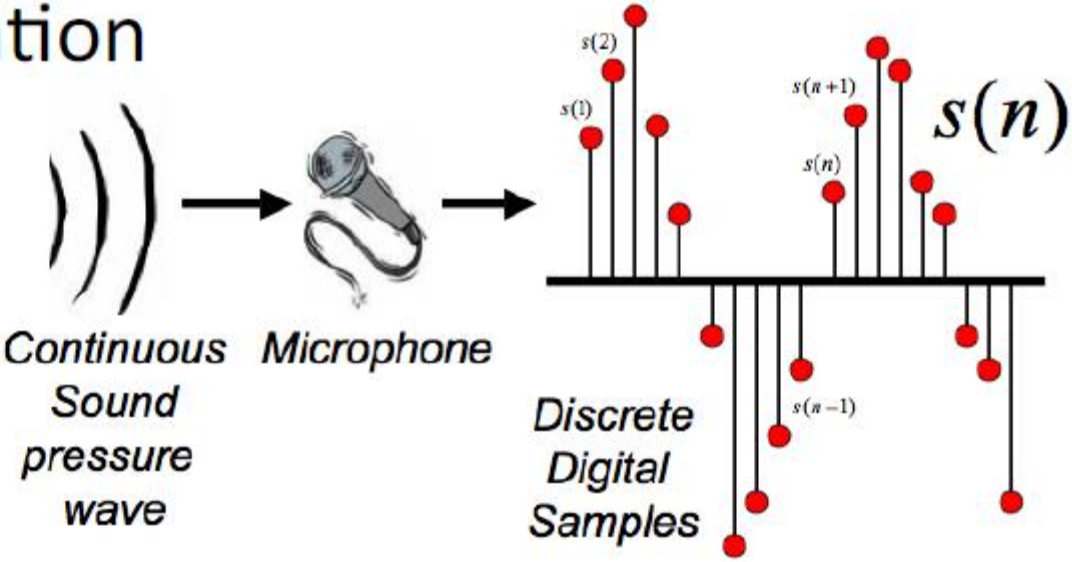


Equation of this wave?



Digitizing speech sounds

- Analog to Digital Conversion
  - Sampling
  - Quantization

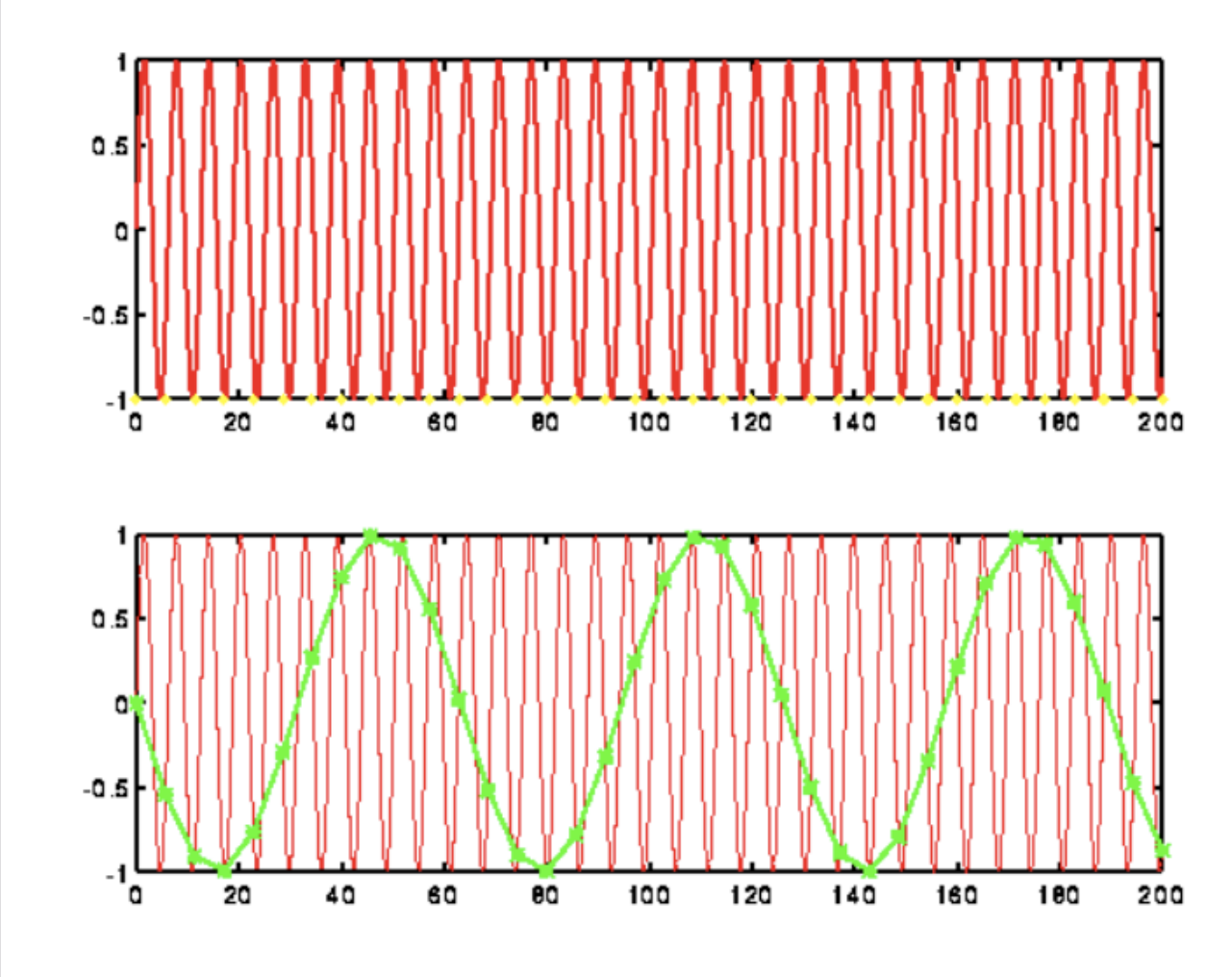


Thanks to Bryan Pellom for this slide!

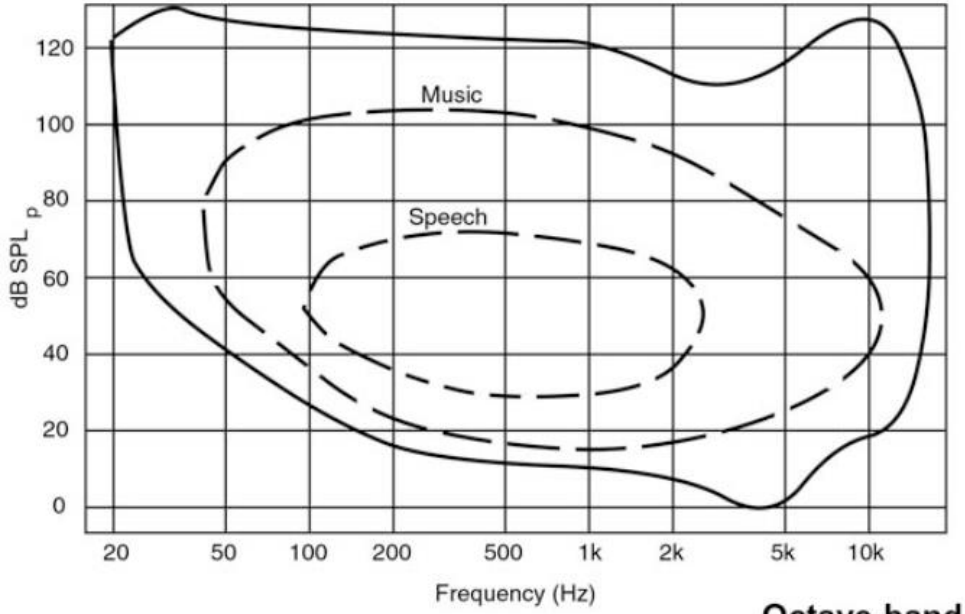
Sampling

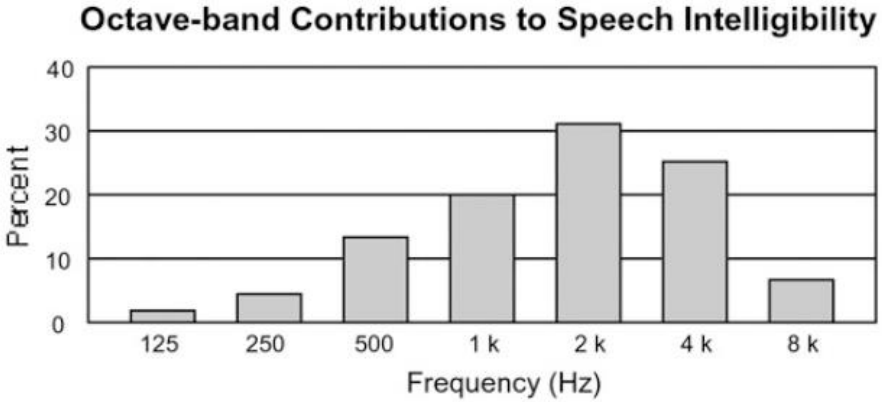
- Nyquist Theorem

Aliasing


1
0.5
0
-0.5
-1
0
20
40
60
80
100
120
140
160
180
200
1
0.5
0
-0.5
-1
0
20
40
60
80
100
120
140
160
180
200

Frequency range of speech





http://www.prosoundweb.com/article/an_examination_of_bandwidth_dynamic_range_and_normal_operating_levels/

Sampling speech sounds

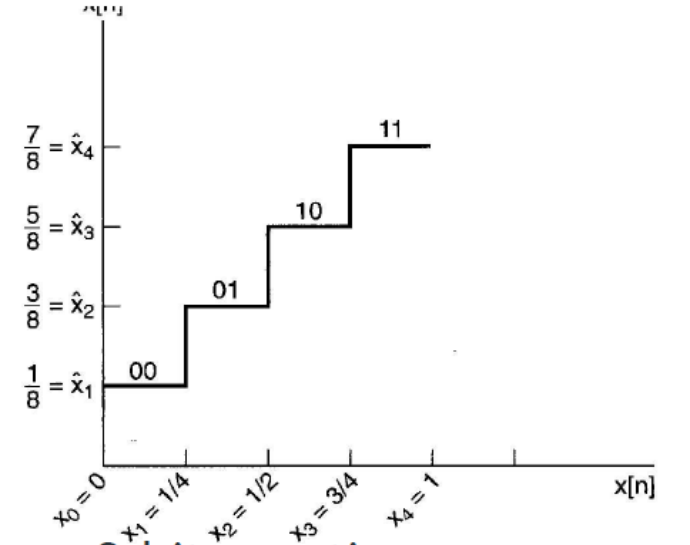
- In practice we use the following sample rates
  - 8 kHz: Telephone
  - 16 kHz for microphones, „wideband“
- Why?
  - Human speech < 10 kHz
  - Telephone is filtered at 4 kHz (300 Hz to 3.4 kHz), so 8 kHz is enough

Quantization

- Usually into 8 bits (256 levels) or 16 bits (65k) levels
- The simplest quantization distributes the real values uniformly among the levels
  - More complicated ones focus on reducing the quantization error

Uniform Quantization

- The decision and reconstruction levels are uniformly spaced.
- In coding it is usually called PCM (Pulse Code Modulation)



2-bit quantizer

Nonuniform Quantization

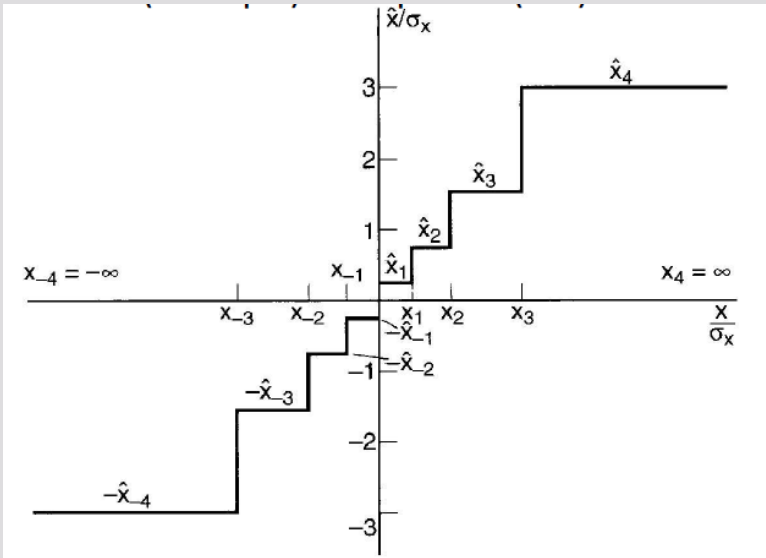
- Reconstruction and decision levels do not have equal spacing.
- Low level signals are more common than high level ones, thus we make quantization steps closer together at the most common signal levels.
- In coding we call this log-PCM.
- Most typical algorithms are A-law (Europe) and μ-law (US)

μ - law

$$y[n] = X_{\max} \frac{\log\left[1+\mu \frac{|x[n]|}{X_{\max}}\right]}{\log[1+\mu]} \operatorname{sign}\{x[n]\}$$

A - law

$$y[n] = X_{\max} \frac{1+\log\left[\frac{A|x[n]|}{X_{\max}}\right]}{1+\log A} \operatorname{sign}\{x[n]\}$$



3-bit nonuniform quantizer

Characteristics of speech signals

- Periodic: voiced
- Aperiodic: unvoiced

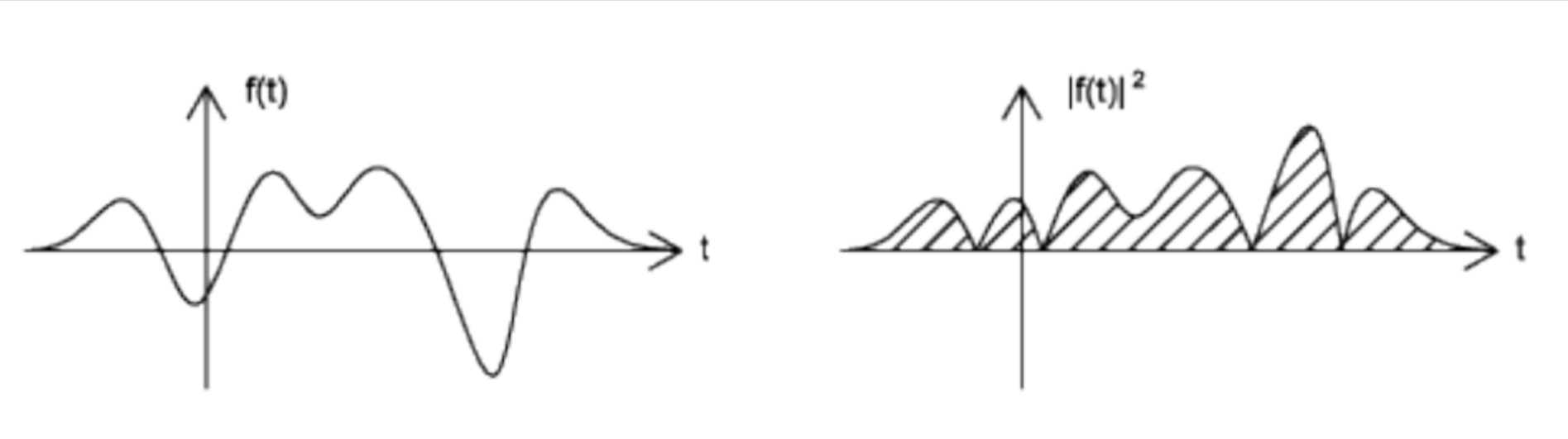
Properties:

- Intensity
- Frequency
- Timbre

Amplitude/energy of a signal

- Signal energy (Joules)

$$E_S = \sum_{i=1}^{N} x[i]^2$$



Power of a signal

- Energy per unit of time. What we call power is usually the „average power“ defined as:

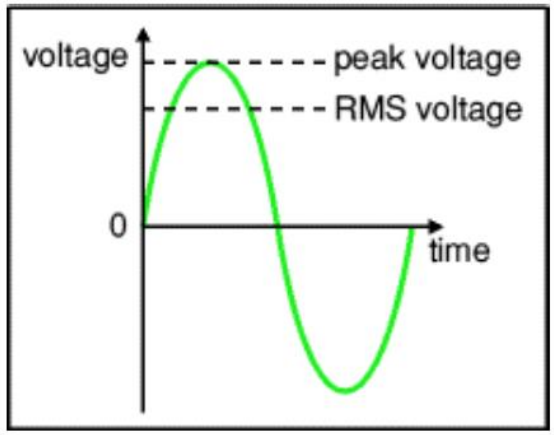
$$P_{avg} = \frac{1}{N}\sum_{i=1}^{N} x[i]^2$$

- Ist dimension is Energy/time (Joules/second) = Watts
- We divide the signal into „windows“ of N samples

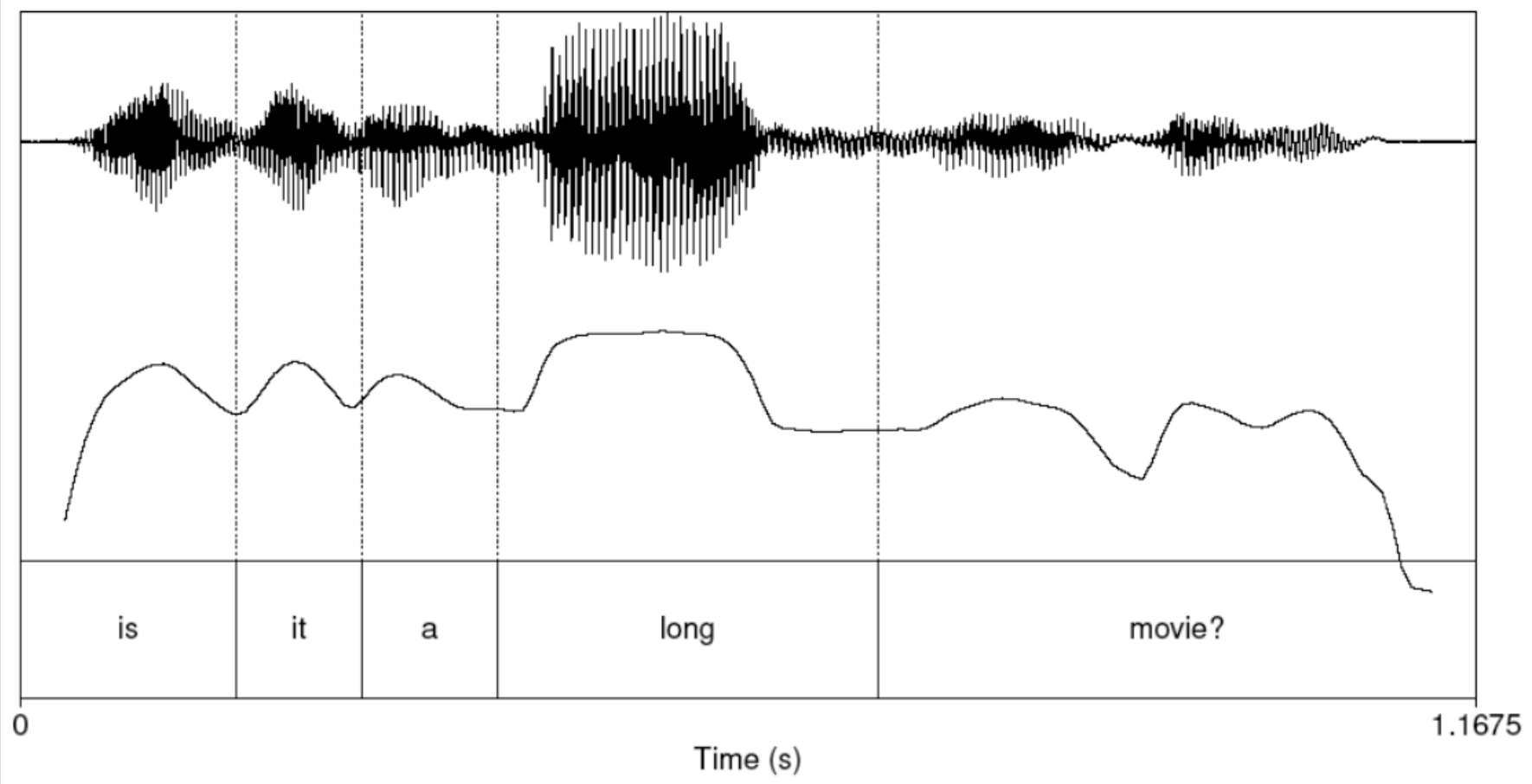
RMS value

- We usually take the root mean square (RMS) value

$$A_{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x[i]^2}$$



Plot of RMS



Sound Pressure Level

- Sound pressure level (SPL) $P_0 = 2^{-5}$Pascals

$SPL = 20log_{10}\frac{p}{p_0}$, p is pressure

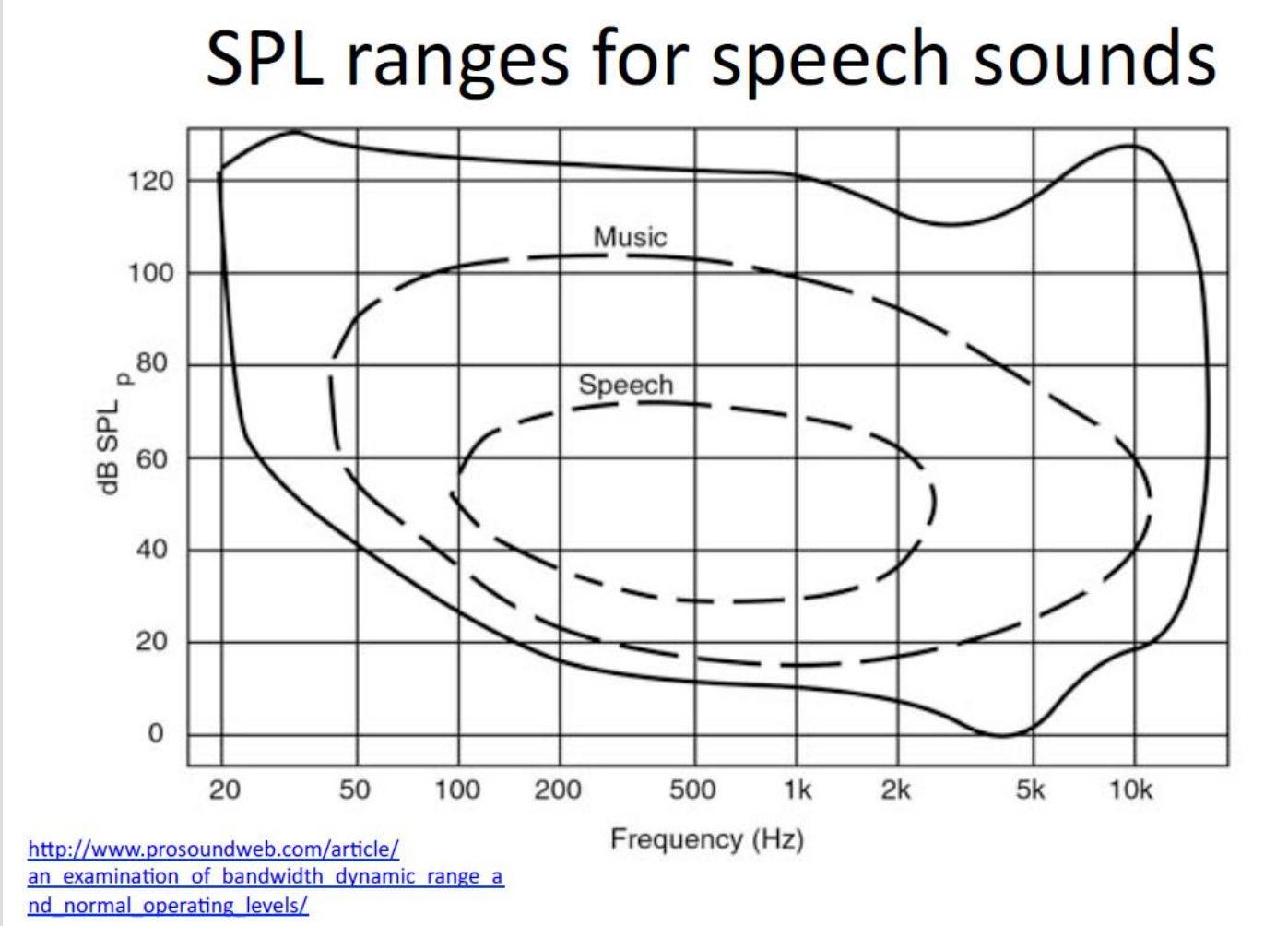
- Sound power level (PWL) and the sound intensity level (IL)

$PWL = 10log_{10}\frac{P}{P_0}$ $IL = 10log_{10}\frac{I}{I_0}$,$P$ is power

- Where $P_0 = 10^{-12}$Watts and $I_0 = 10^{-12}$Watts/m^2

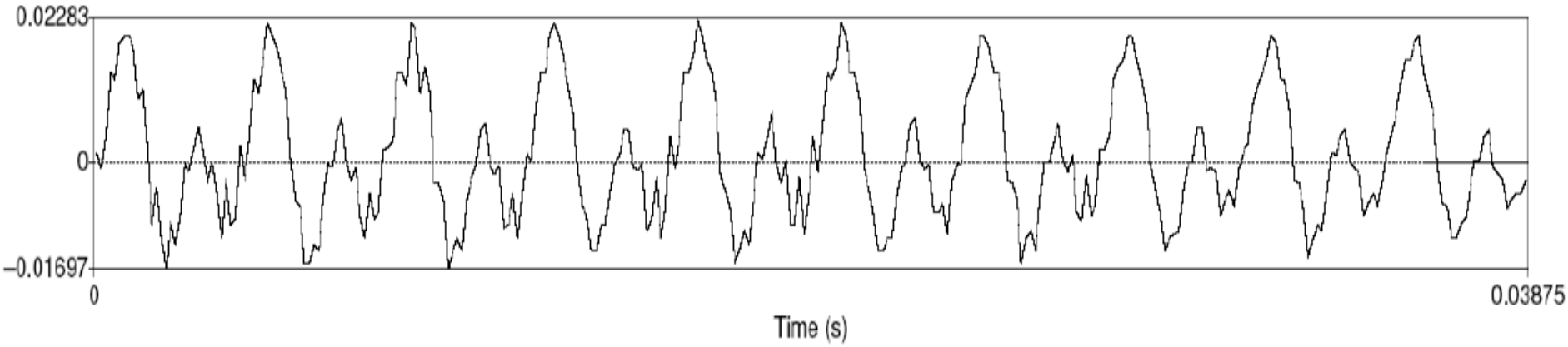
http://www.sengpielaudio.com/calculator-soundpower.htm

SPL ranges for speech sounds



Fundamental Frequency

- Waveform of a vowel (voiced signal)
- Fundamental frequency = 1/T
  - Vowel has 10 reps in .03875 secs
    → freq is 10/.03875 = 258 Hz



Ranges of f0 in speech

- Male: 85-180 Hz
- Female: 165-255 Hz

Pitch

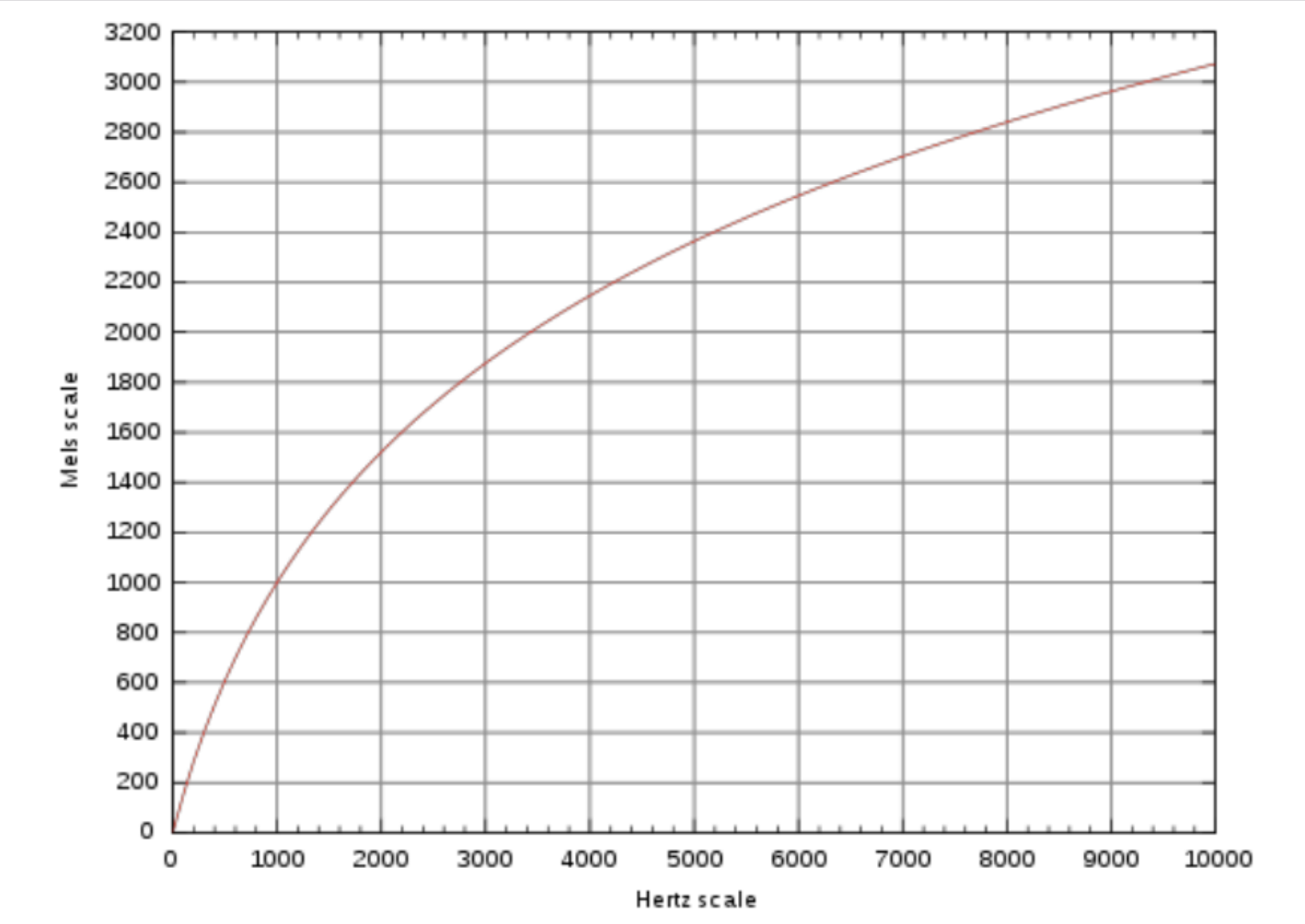
- Pitch: perceived fundamental frequency
- Non linear relationship:
  - Human pitch perception is most accurate between 100 Hz and 100 Hz.
    - Linear in this range: At F_{01}=200 Hz, if Pitch2=Pitch1/2 then F_{02} ≈100 Hz
  - Logarithmic above 1000 Hz: At $F_{01} = 5$ kHz if Pitch2 = Pitch1/2 then $F_{02} \approx 2\ kHz$
- Still, in the literature many times F0 and pitch are treated as the same

Pitch vs. F0 modeling

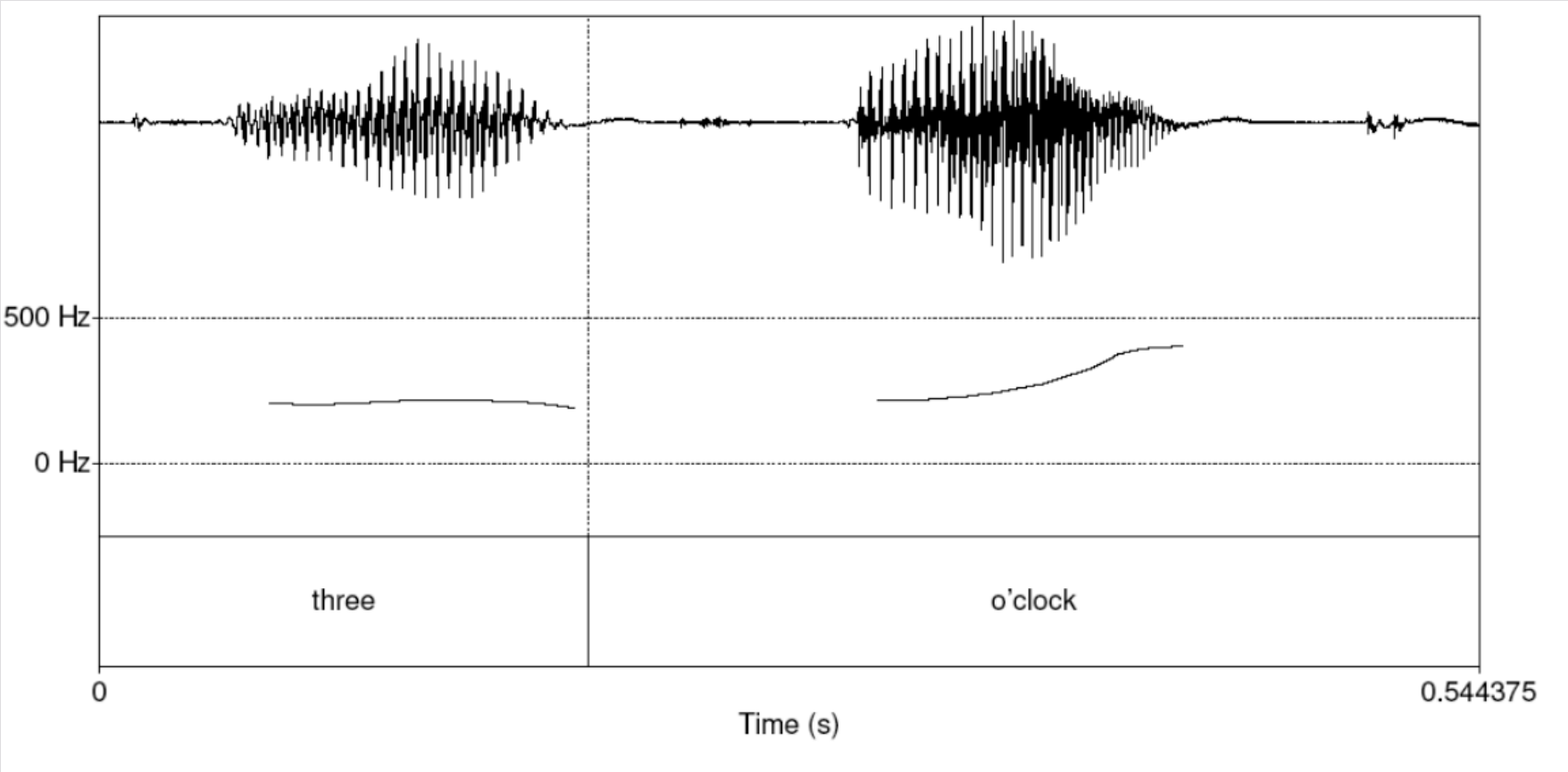
- Mel scale in one model of F0-pitch mapping

$$m = 2595 log_{10}\left(\frac{f}{700} + 1\right)$$
$$= 1127 log_e\left(\frac{f}{700} + 1\right)$$

Pitch mels vs Hz

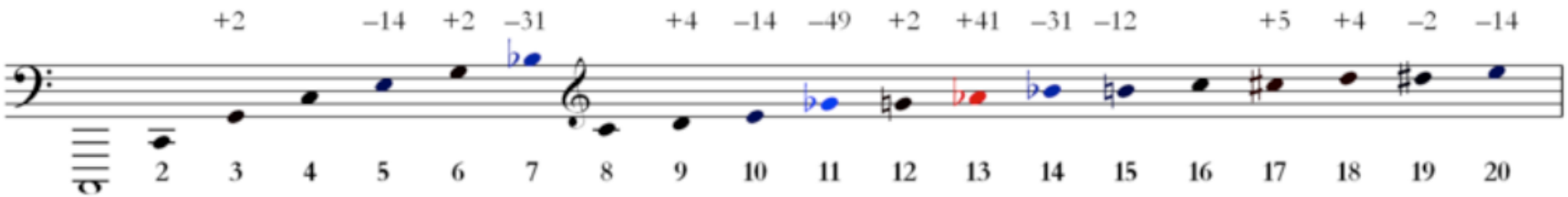


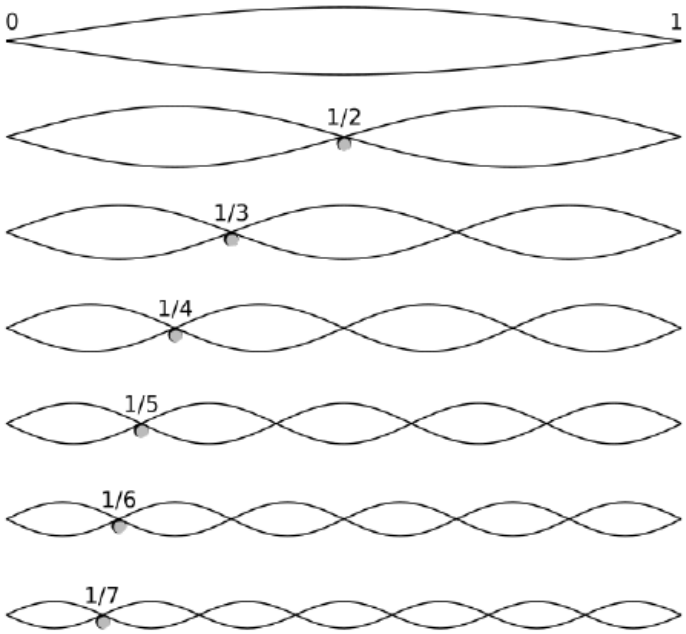
F0 tracking: prosody



- F0 can be computed using several techniques, and using tools like PRAAT

Harmonic components (voiced sounds)





Fundamental Frequency = first harmonic

2nd harmonic

3rd harmonic

4th harmonic

5th harmonic

6th harmonic

7th harmonic

Spectrum

- Representation of energy at the different frequency components:
  - Harmonic components
  - Partials (unvoiced sounds or components of voiced sounds)

Fourier transform anlysis

- Fourier analysis: any wave can be represented as the (infinite) sum of sine waves of different frequencies (amplitude, phase)
- For continuous signals

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi t f} dt$$

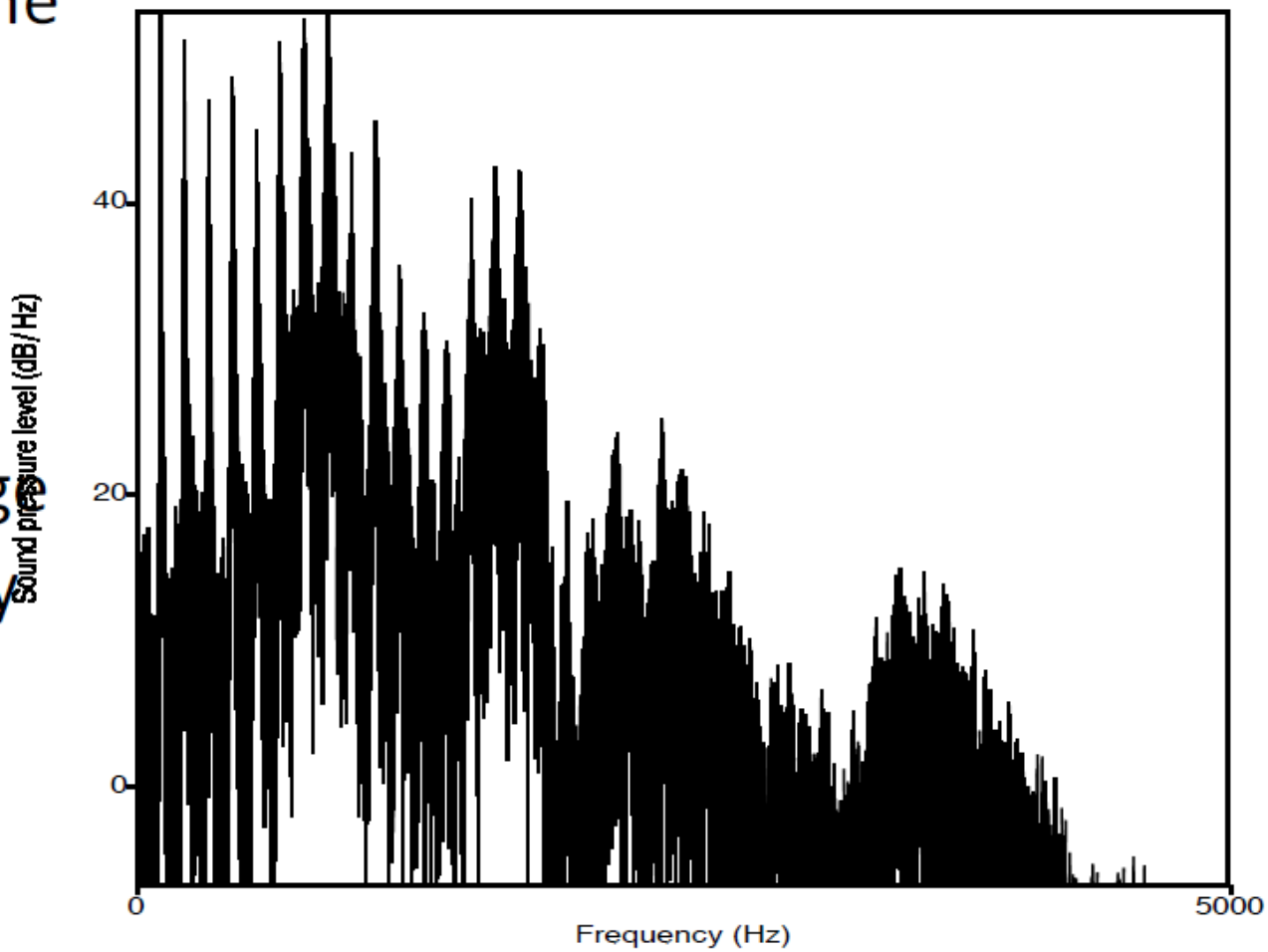
- For discrete signals

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i\frac{2\pi}{N}kn} \quad k = 0,\ldots,N-1$$

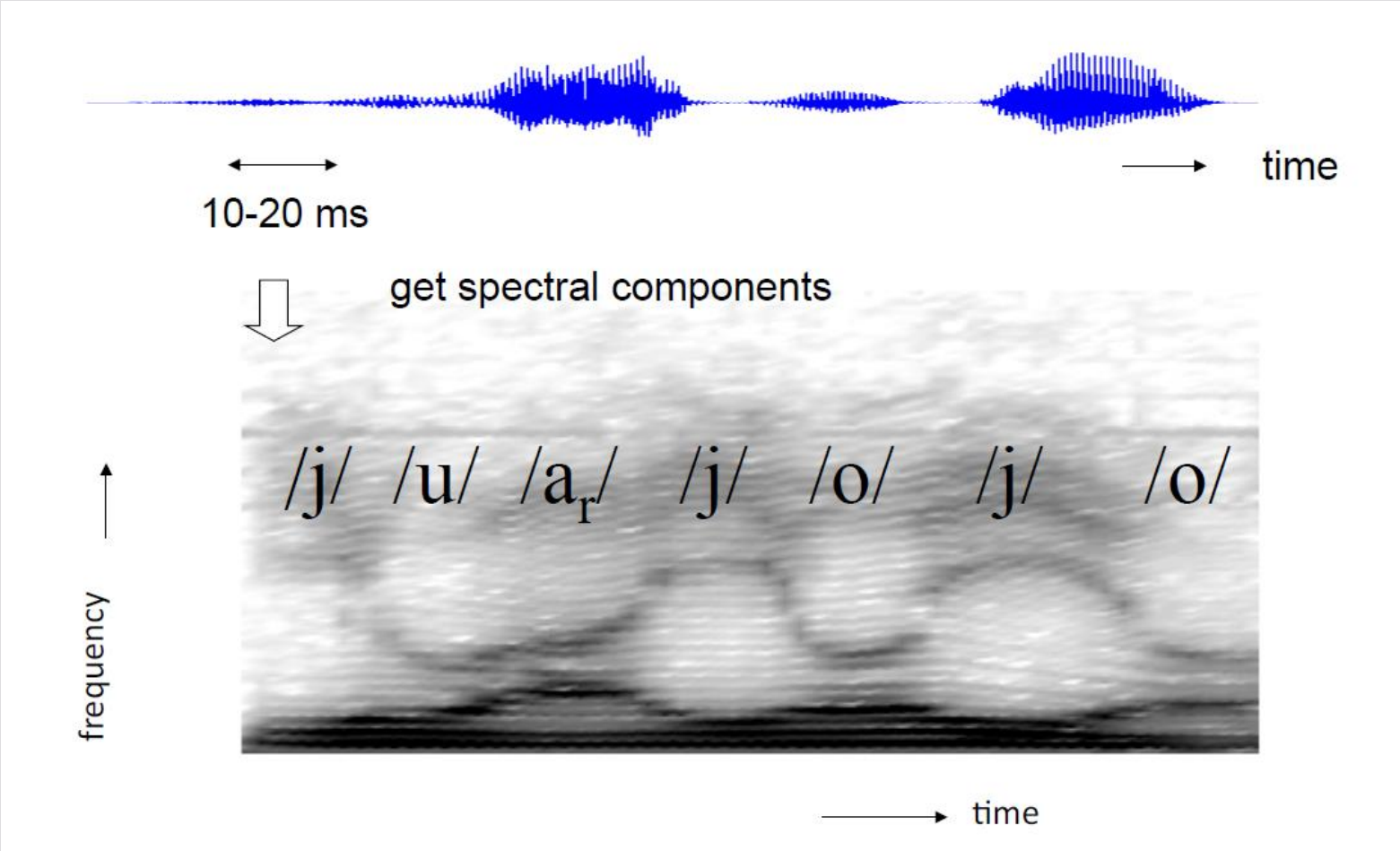
- When N is finite (and relatively short) we call the resulting signal the short term spectrum

Spectrum Example

- Spectrum of one instant in an actual sound wave: many components across the frequency range
- Each frequency component of the wave is separated

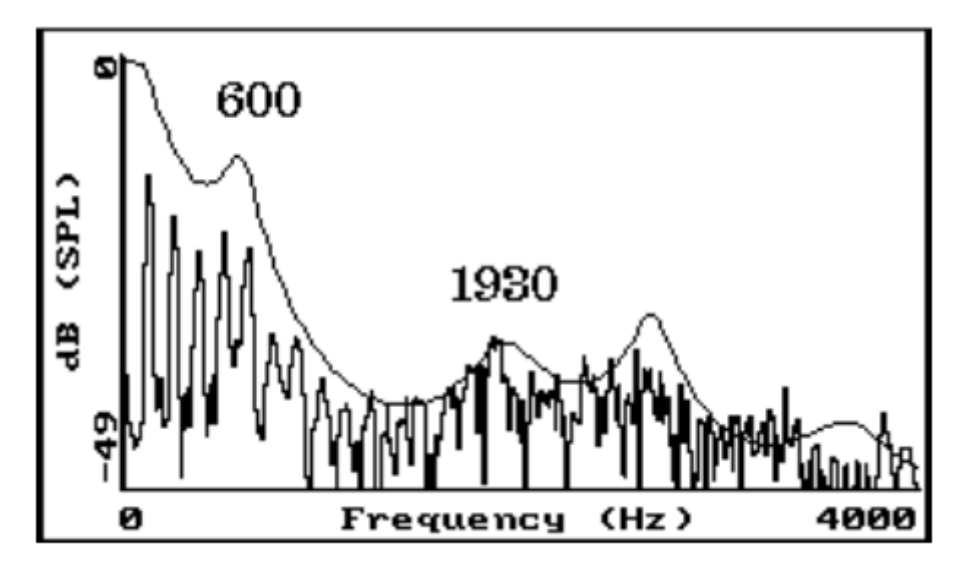


Short-term spectrum of a speech signal

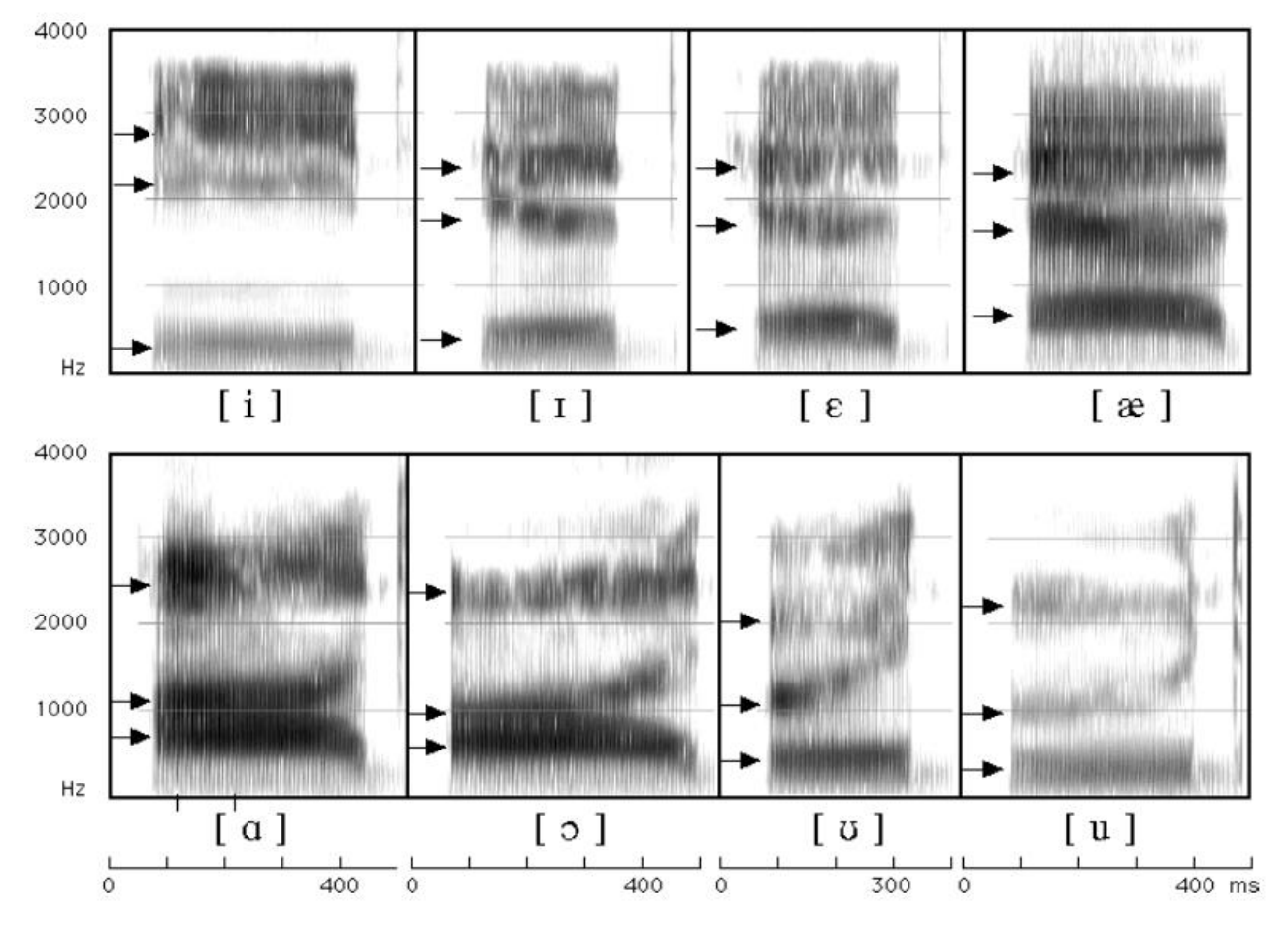


Formants

- Formants are defined as the spectral peaks of the sound spectrum envelope
- Formants are independent of the F0 frequency, as they are defined over the envelope of the spectrum
- They are created by the pass of the sound through the vocal tract

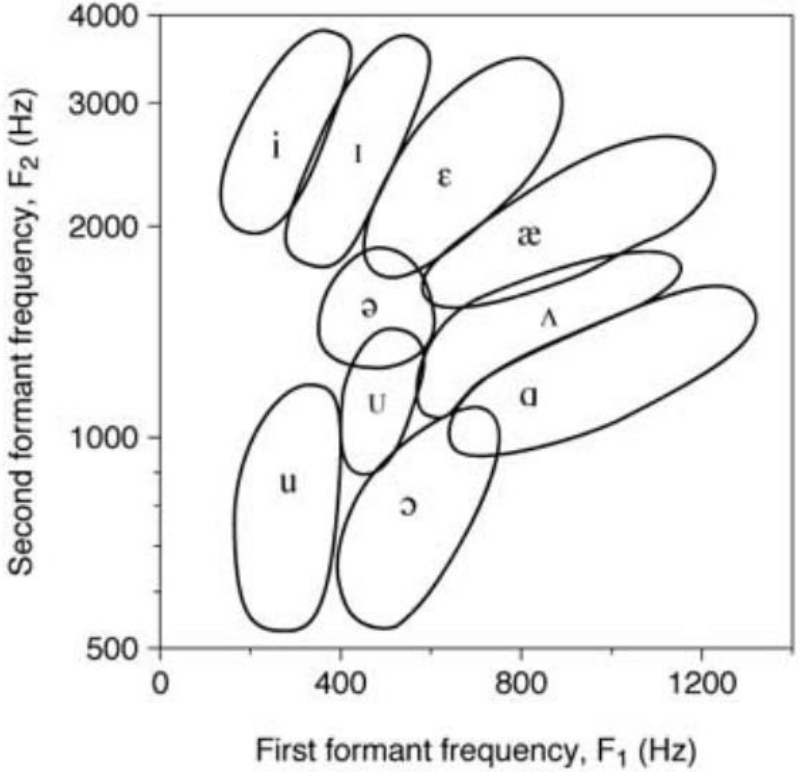


Seeing formants: the spectrogram



Formants in the vowels (American English)

- http://www.uiowa.edu/~acadtech/phonetics/english/frameset.html



Example