

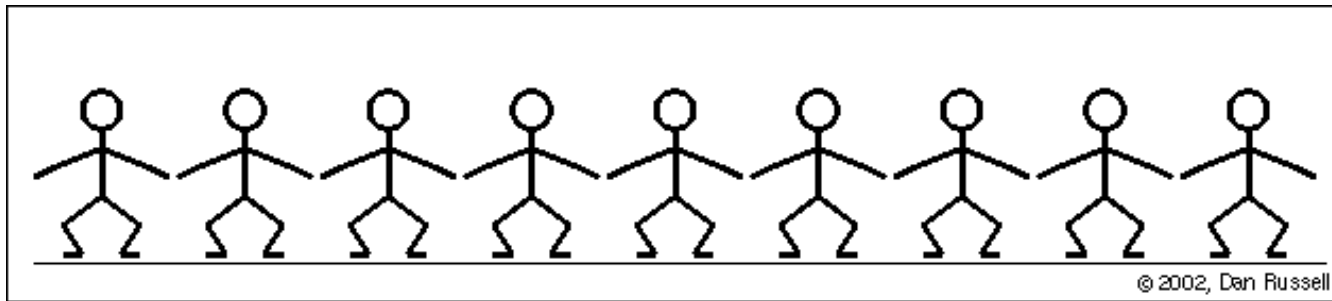
1. Introduction to acoustics

Block 1

(Many slides come from materials from Dan Jurafsky)

Acoustic Phonetics

- **Sound Waves:** "a disturbance or variation that transfers energy progressively from point to point in a medium and that may take the form of an elastic deformation or of a variation of pressure, electric or magnetic intensity, electric potential, or temperature.", Webster's dictionary

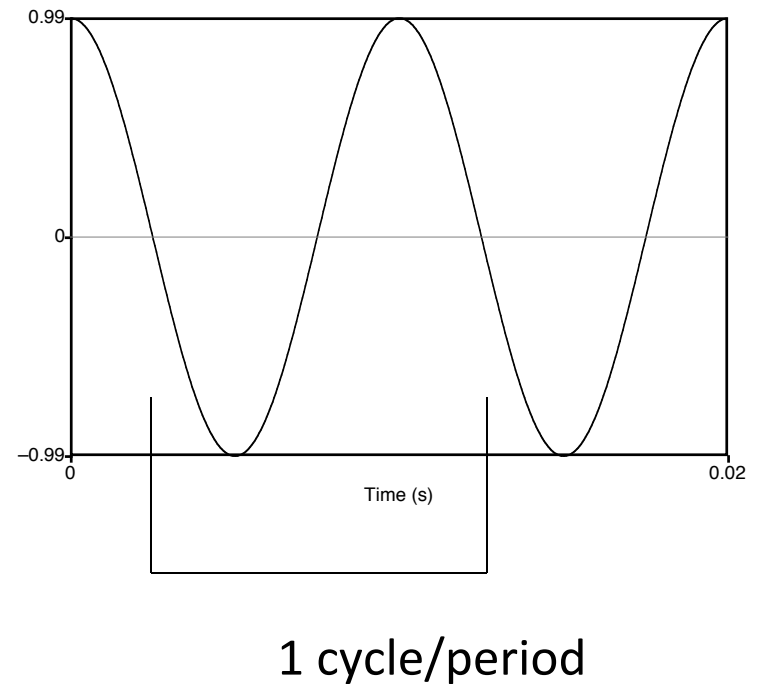


©2002, Dan Russell

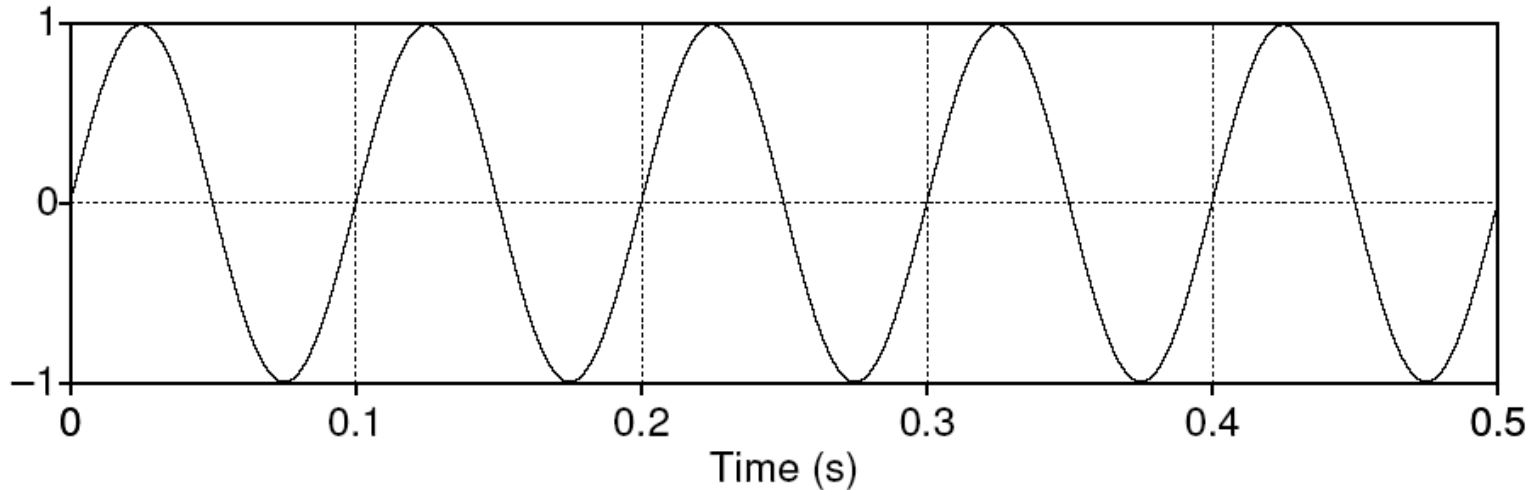
Animation courtesy of Dr. Dan Russell, Kettering University

Simple Period Waves (sine waves)

- Characterized by:
 - period: T
 - amplitude A
 - phase ϕ
- Fundamental frequency in cycles per second, or Hz
 - $F_0 = 1/T$



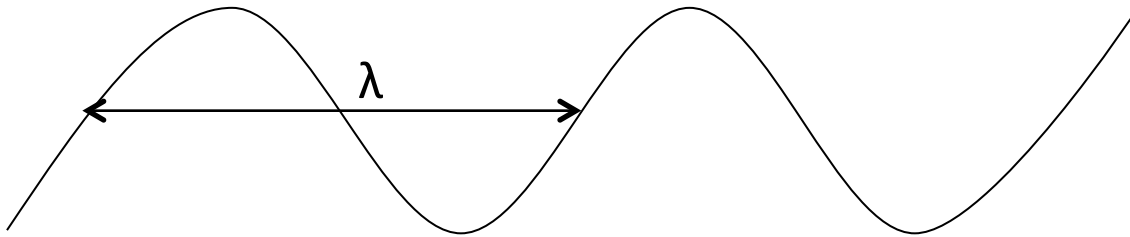
Simple periodic waves



- Computing the frequency of a wave:
 - 5 cycles in .5 seconds = 10 cycles/second = 10 Hz
- Amplitude:
 - 1
- Equation:
 - $Y = A \sin(2\pi ft)$

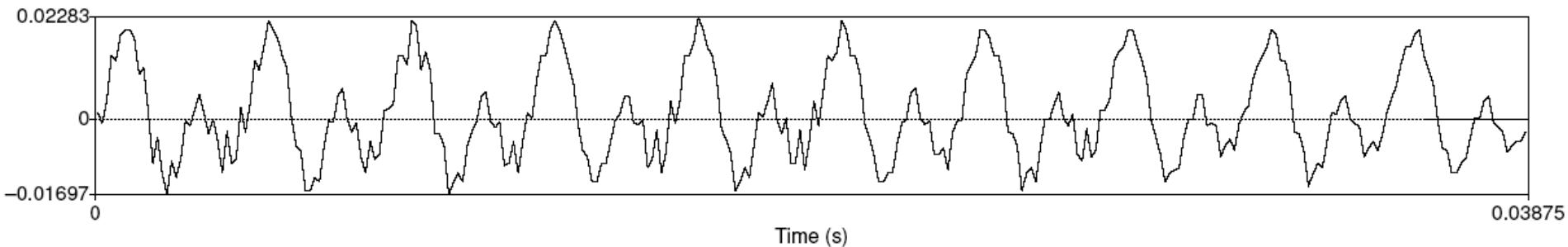
(more) Basic facts about sound waves

- $f = c/\lambda$
- Where c = speed of sound, and λ = wave length (longitudud de onda, in meters)



- $c=3440$ cm/s (≈ 350 m/s) at 21 degrees Celsius at sea level
- Example: with $\lambda=10$ m, frequency $f=35$ Hz

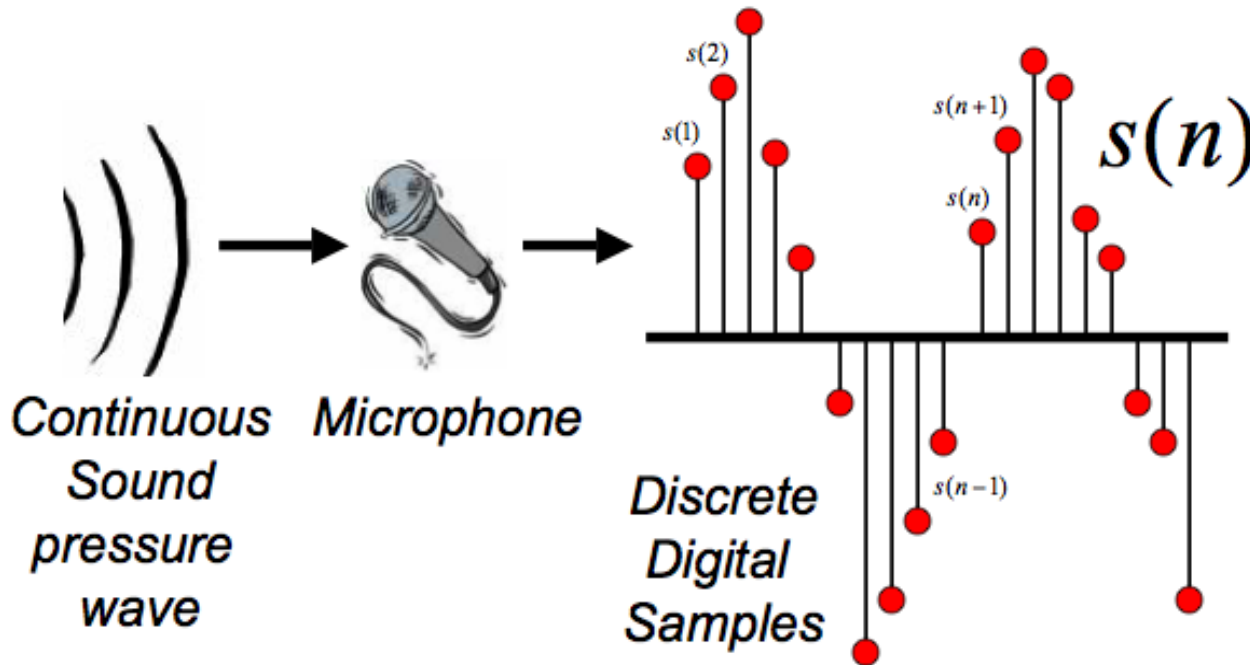
Speech sound waves



- A little piece from the waveform of the vowel [iy]
- Y axis:
 - Amplitude = amount of air pressure at that time point
 - Positive is compression
 - Zero is normal air pressure,
 - negative is rarefaction (expansion)
- X axis: time.

Digitizing Speech (Analog-to-digital conversion)

- Two steps
 - Sampling
 - Quantization



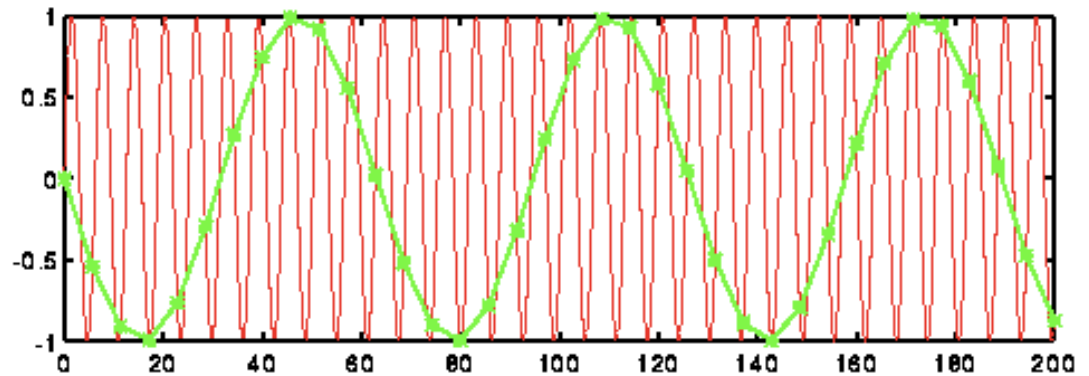
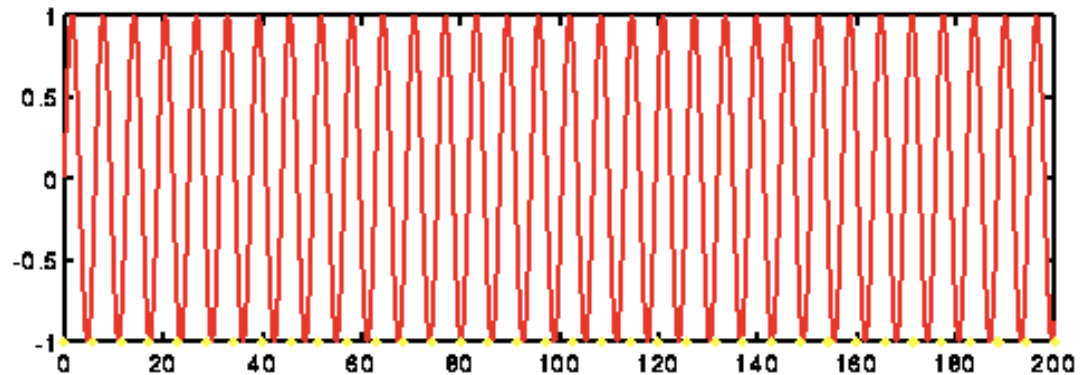
Sampling

- Measuring amplitude of a signal at time t
- The sample rate needs to have at least two samples for each cycle/period
 - One for the positive, and one for the negative half of each cycle
 - More than two samples per cycle is ok; Less than two samples will cause frequencies to be missed
- So the maximum frequency that can be measured and stored is half the sampling rate.
 - Nyquist Theorem: $F_{\text{niq}} = F_s / 2$
 - Any existing frequency above F_{niq} will cause aliasing.

Sampling

Original signal in red:

If measure at green dots, will see a lower frequency wave and miss the correct higher frequency one!



Sampling

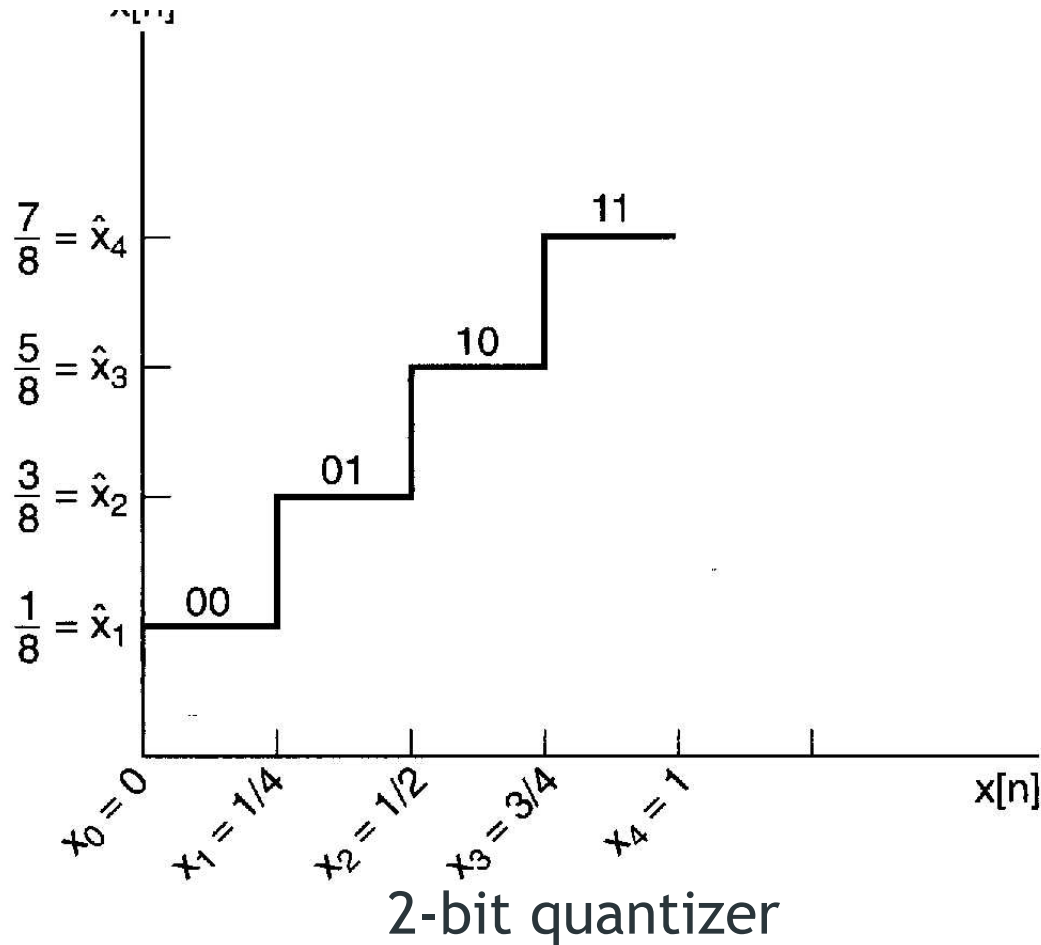
- In practice we use the following sample rates
 - 16,000 Hz (samples/sec), for microphones, “wideband”
 - 8,000 Hz (samples/sec) Telephone
- Why?
 - Need at least 2 samples per cycle
 - Max measurable frequency is half the sampling rate
 - Human speech < 10KHz, so need max 20K
 - Telephone is filtered at 4K (300Hz to 3.4KHz), so 8K is enough.

Quantization

- Representation of every sampled real value as an integer.
- Usually into 8 bits (256 levels) or 16 bits (65K) levels, although any number of bits can be used.
- 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)



Nonuniform Quantization

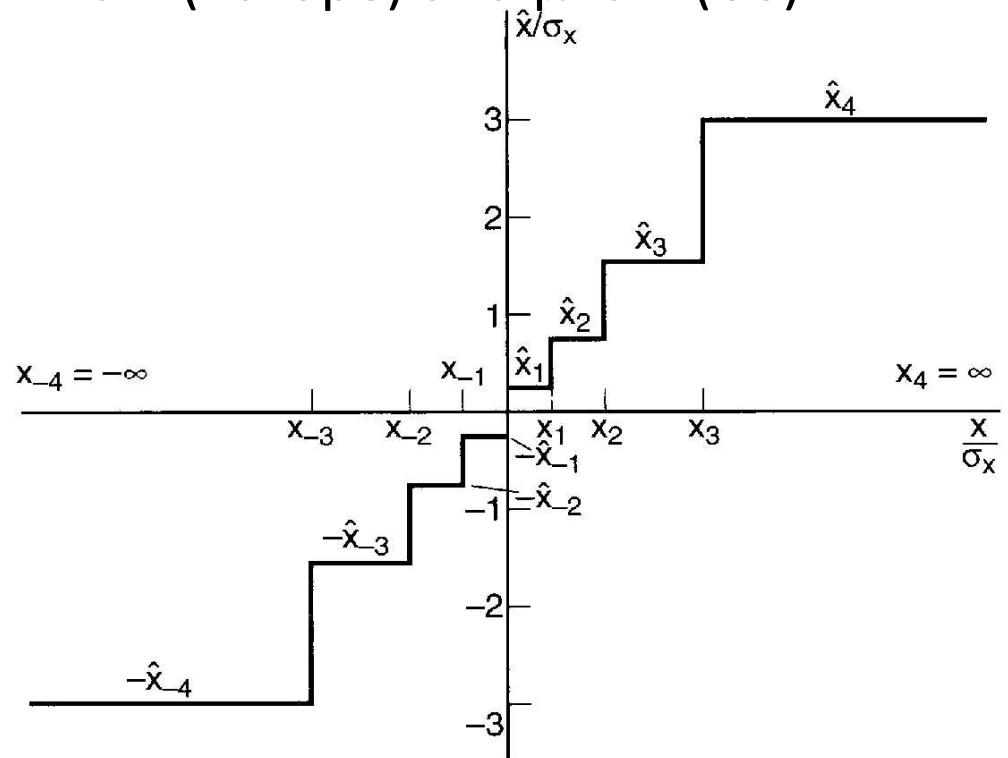
- Reconstruction and decision levels do not have equal spacing.
- Low level signals are more common than high level ones, thus we make quantization steps closer together at the most common signal levels.
- In coding we call this log-PCM
- Most typical algorithms are A-law (Europe) and μ -law (US)

μ - law

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

A - law

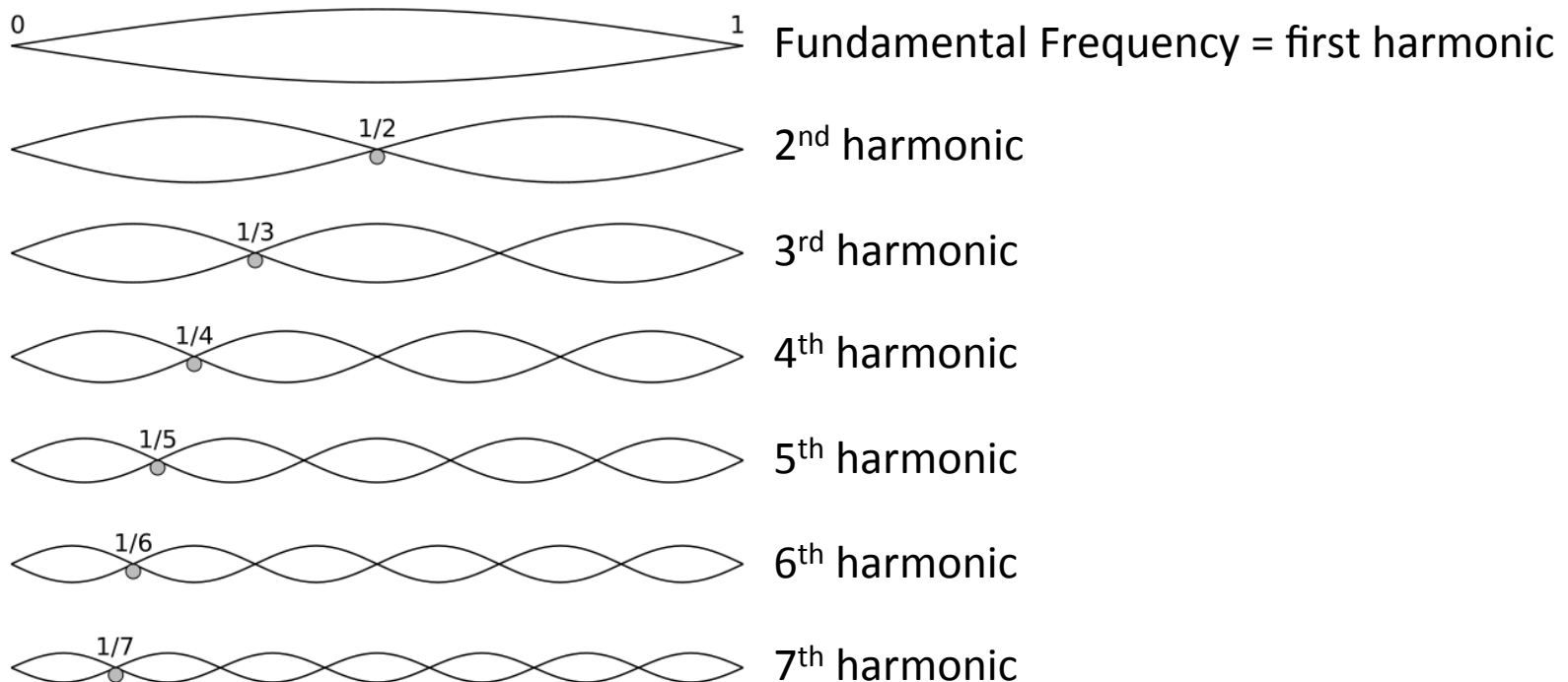
$$y[n] = X_{\max} \frac{1 + \log \left[\frac{A|x[n]|}{X_{\max}} \right]}{1 + \log A} \text{sign}\{x[n]\}$$



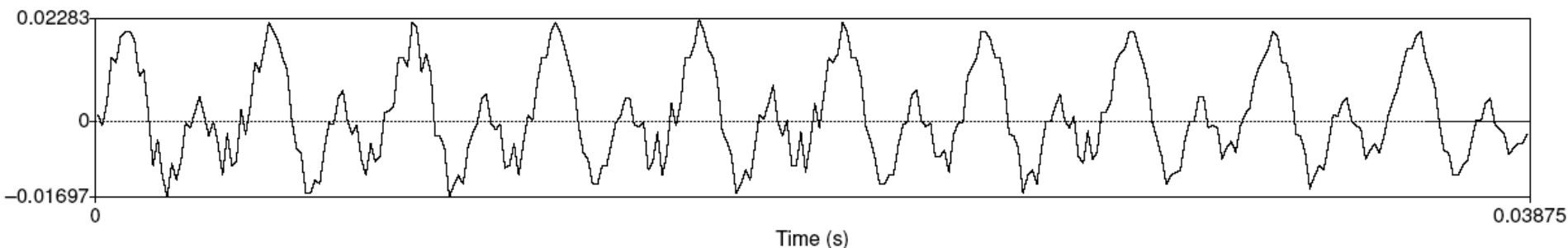
3-bit nonuniform quantizer

Fundamental frequency

- The fundamental frequency (or F_0) is the lowest frequency of a periodic (voiced) waveform, produced by any particular instrument (our vocal cords are like a “complicated” instrument)
- It is also called the first harmonic, in comparison with its integer multiples called second, third, etc. harmonics



Fundamental frequency



In speech, see for example the waveform of the vowel [iy]

- The fundamental frequency could be computed as the minimum number of repetitions/second of the wave:
 - Above vowel has 10 reps in .03875 secs -> freq. is $10/.03875 = 258$ Hz
- This is the speed that vocal folds move, hence voicing
- Each peak corresponds to an opening of the vocal folds

Pitch

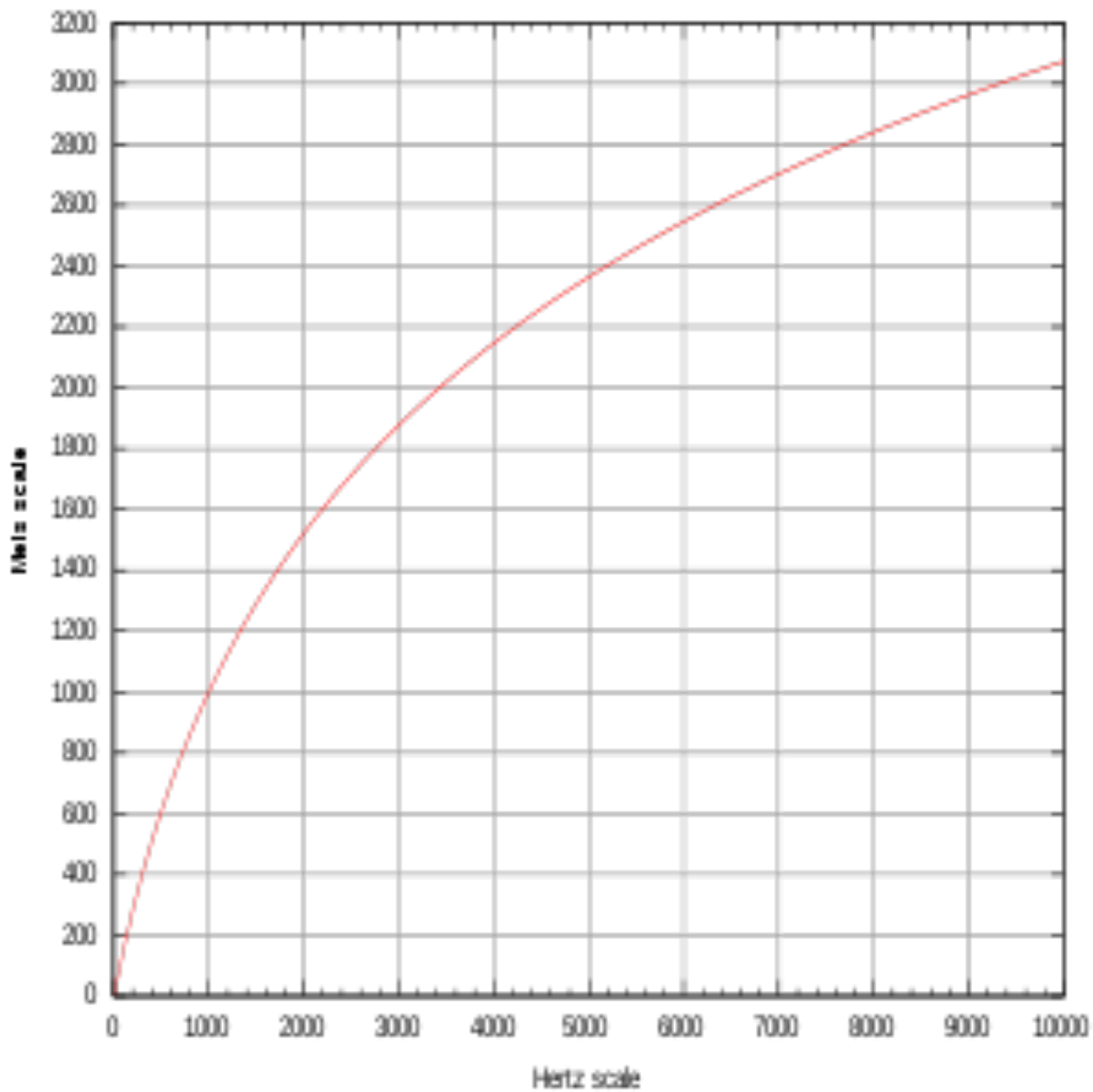
- Pitch is defined as the **perceived** fundamental frequency of a sound
- F0 and pitch are different concepts:
 - F0 corresponds to the physically measurable minimum frequency
 - Pitch corresponds to the minimum perceivable frequency
- The relationship between pitch and F0 is not linear
 - human pitch perception is most accurate between 100Hz and 1000Hz.
 - Linear in this range: At $F0_1=200\text{Hz}$, if $\text{Pitch}_2=\text{Pitch}_1/2$ then $F0_2\approx 100\text{Hz}$
 - Logarithmic above 1000Hz: At $F0_1=5\text{KHz}$ if $\text{Pitch}_2=\text{Pitch}_1/2$ then $F0_2<2\text{KHz}$
- Still, in the literature many times F0 and pitch are treated as the same

Pitch vs. F0 modeling

- Mel scale is one model of this F0-pitch mapping
 - A mel is a unit of pitch defined so that pairs of sounds which are perceptually equidistant in pitch are separated by an equal number of mels

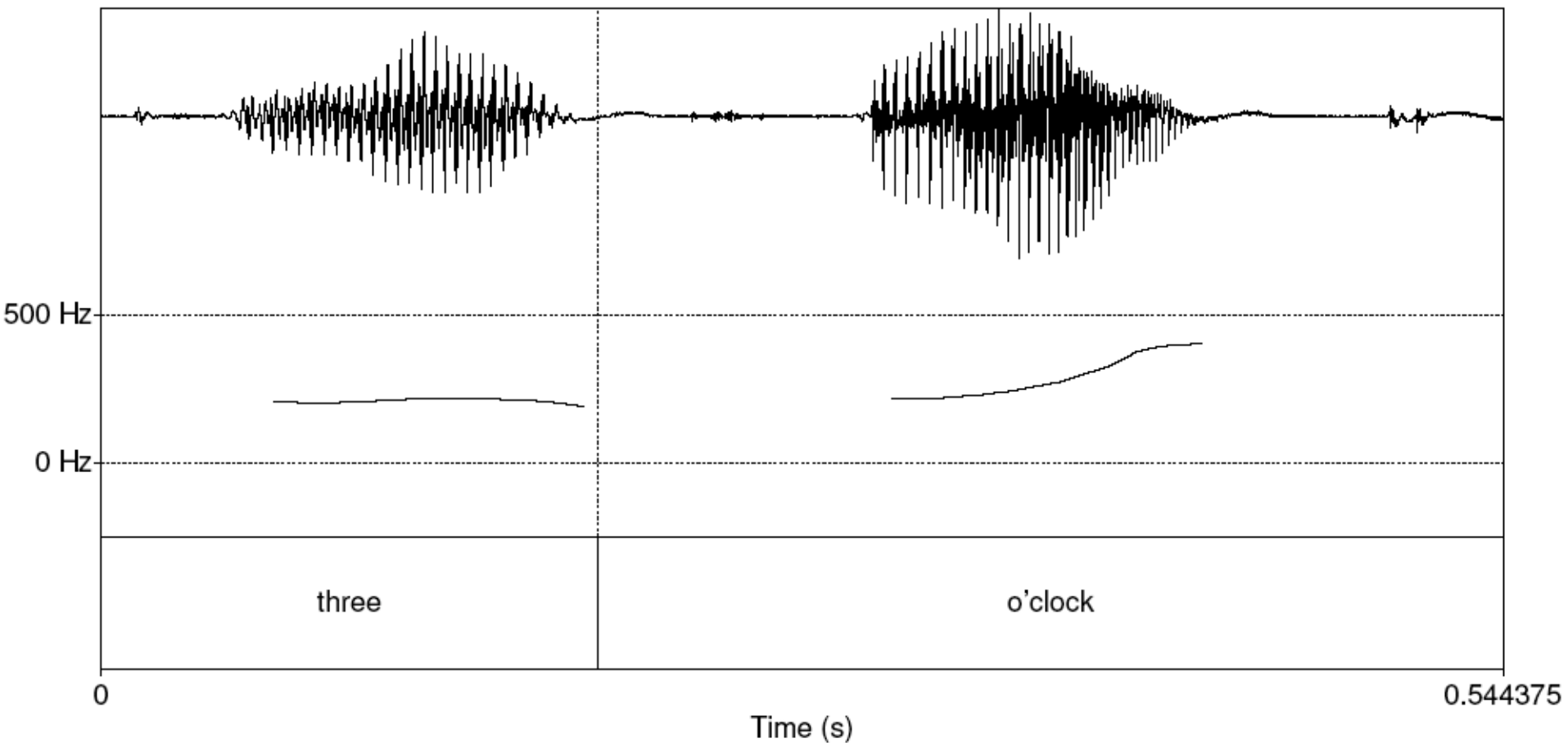
$$m = 2595 \log_{10} \left(\frac{f}{700} + 1 \right) = 1127 \log_e \left(\frac{f}{700} + 1 \right)$$

Pitch vs. F0 modeling





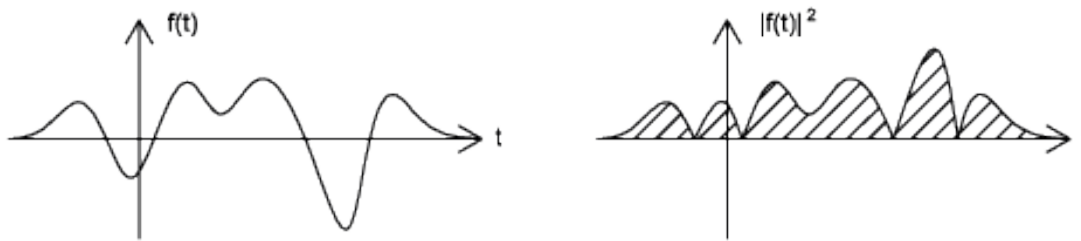
F0 tracking



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

Amplitude of a signal

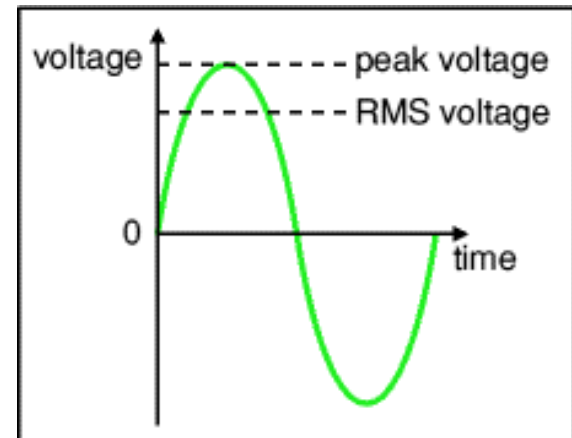
- We need a way to talk about the amplitude of a region of a signal over time
- The signal energy is defined as



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

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

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



Power of a signal

- The power of a signal (P) is its energy per unit of time. What we call power is usually the “average power” defined as

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

Its dimension is Energy/time (**Joules/second**) which is called **Watts**

Pressure and intensity

- Sound intensity (I): it is the power per unit of area (in W/m^2). It is a physical measure, difficult to obtain.
- Sound pressure (p): it is the magnitude of the pressure variations that floating particles experience when a waveform propagates.



Measured in Pascals ($\text{Newtons}/\text{m}^2$)

Our ears or a microphone react to sound pressure rather than to intensity. In progressive plane waves Intensity \approx (pressure)²

Sound pressure level

- Sound pressure level (SPL) is the most used metric. It is the ratio between the actual sound pressure and some relative reference value. It is measured in decibels (db). The most common reference used is the threshold of human hearing $P_0 = 2 \times 10^{-5}$ Pascals

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

- Similarly, we can define the sound power level (PWL) and the sound intensity level

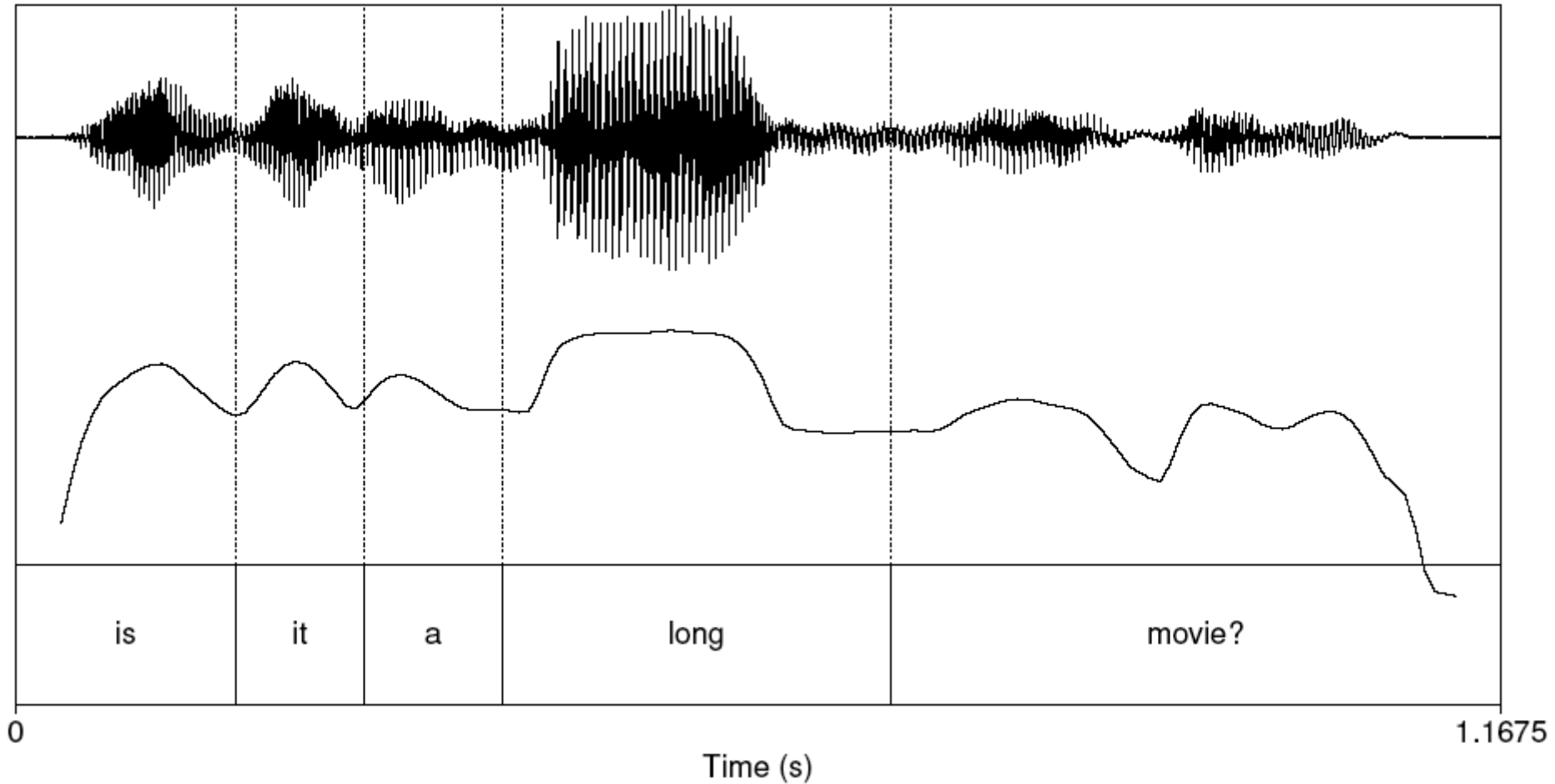
$$PWL = 10 \log_{10} \frac{P}{P_0} \qquad IL = 10 \log_{10} \frac{I}{I_0}$$

Where $P_0 = 10^{-12}$ Watts and $I_0 = 10^{-12}$ Watts/m²

In normal conditions they are all linearly relative to each other

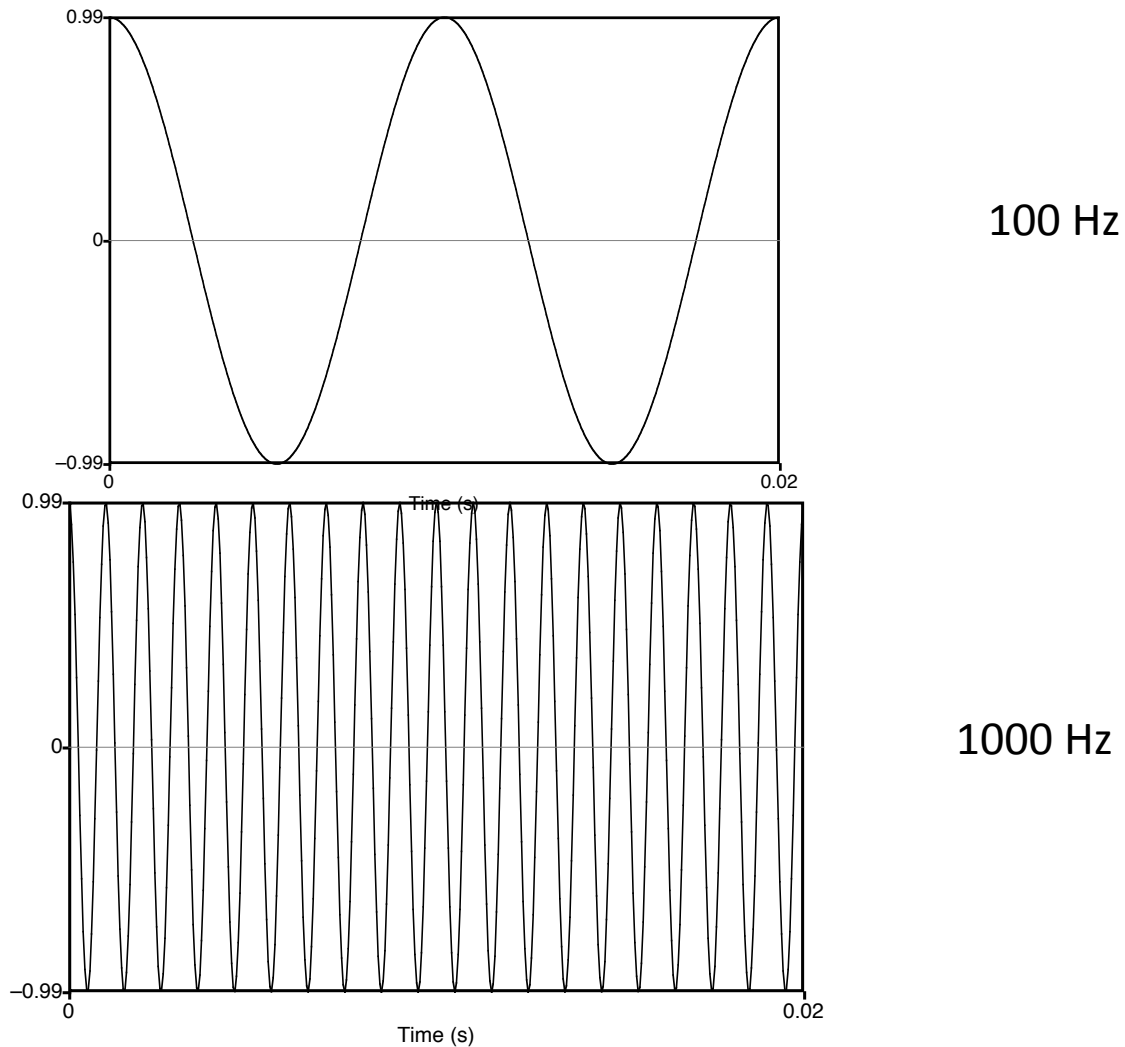


Plot of Intensity



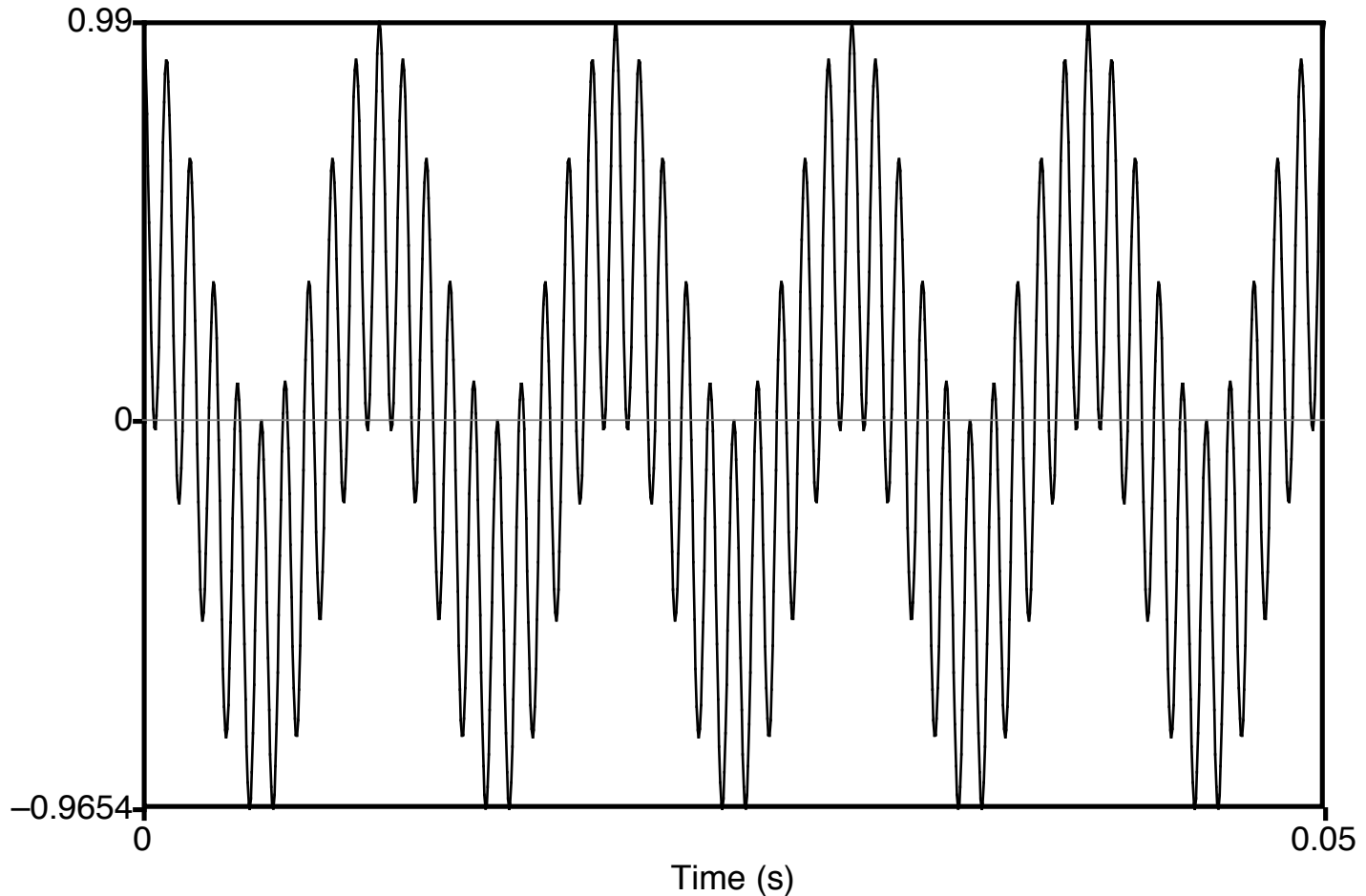
Frequency analysis

- Waves have different frequencies



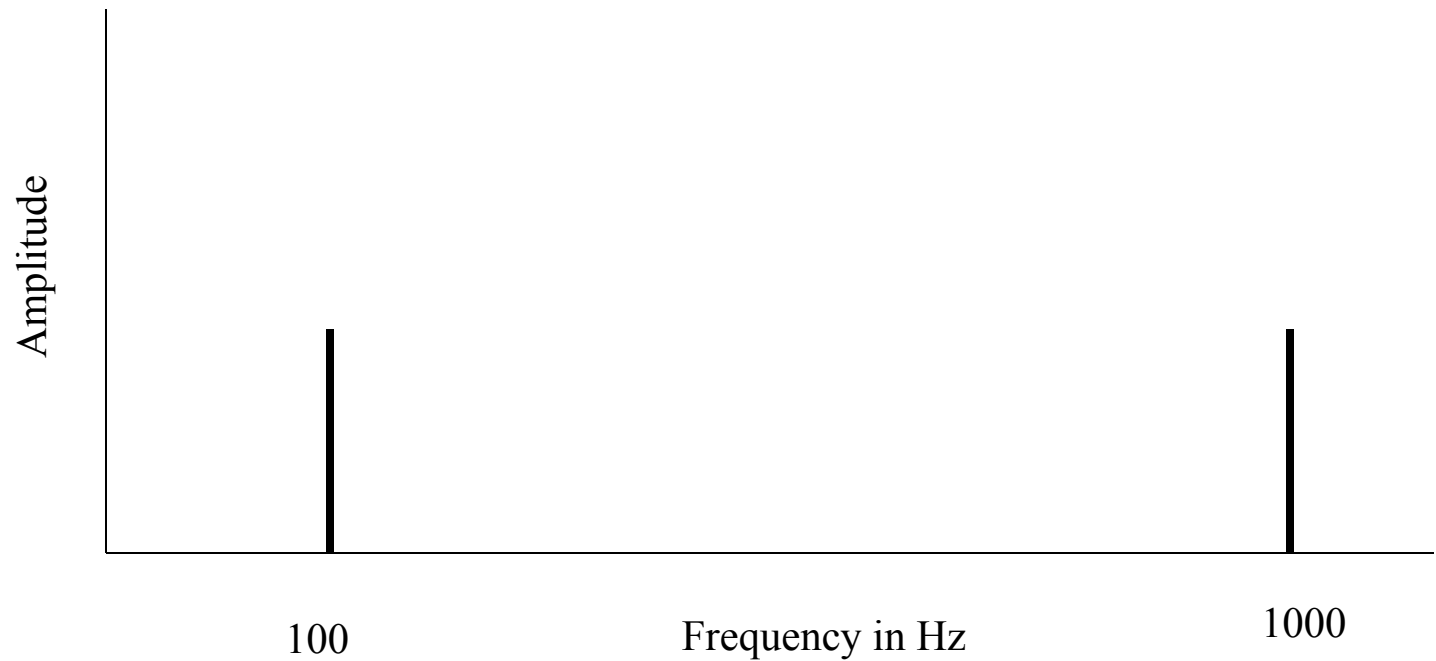
Frequency analysis

- Complex waves: Adding a 100 Hz and 1000 Hz wave together



Spectrum

Frequency components (100 and 1000 Hz) on x-axis



Fourier transform analysis

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

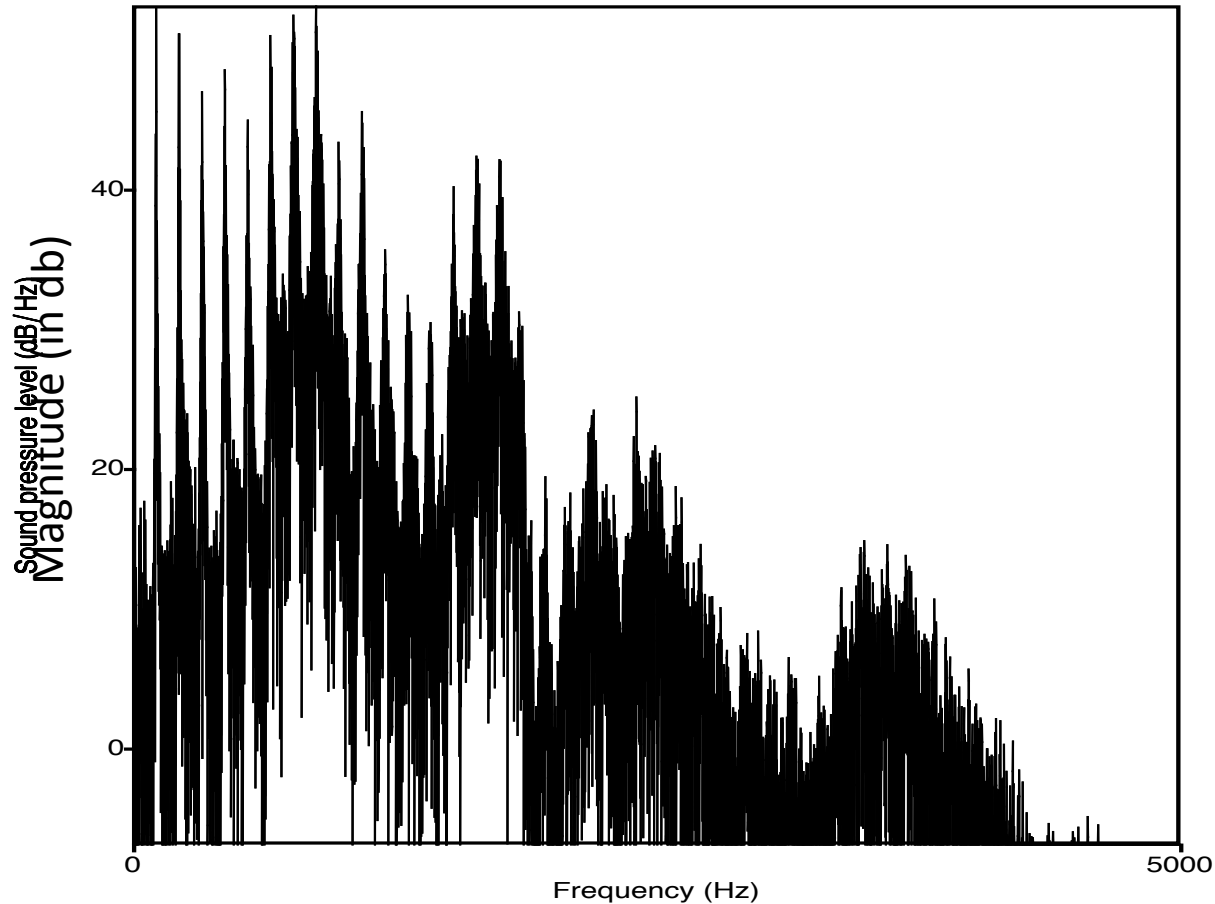
$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx,$$

- For discrete signals:

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

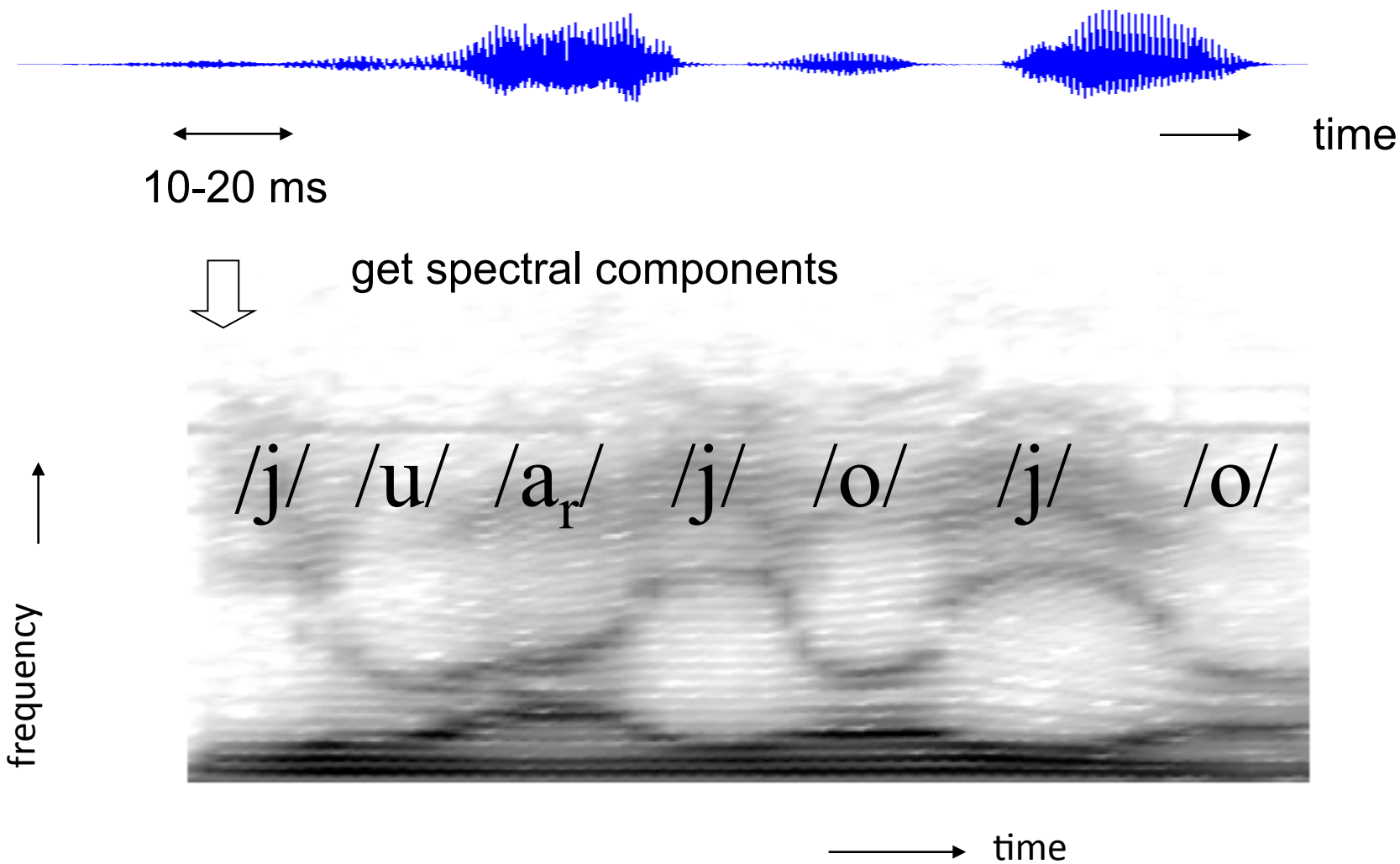
When N is finite (and relatively short) we call the resulting signal the short term spectrum

Spectrum example



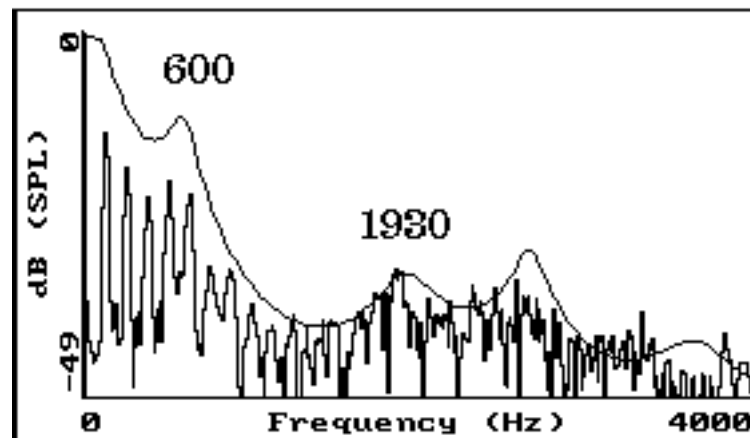
- Spectrum of one instant in an actual soundwave: many components across the frequency range
- Each frequency component of the wave is separated

Short-term Spectrum

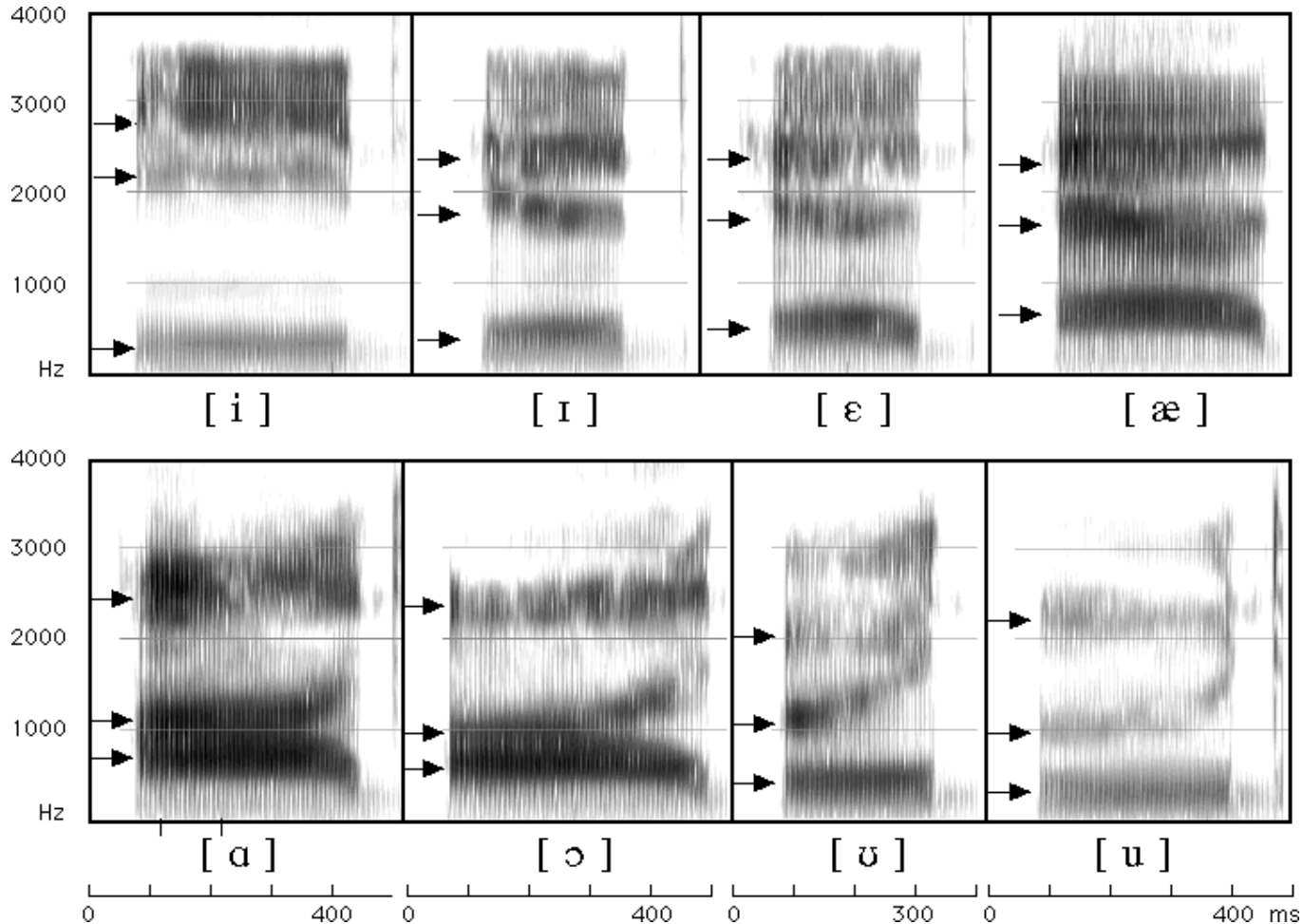


Formants

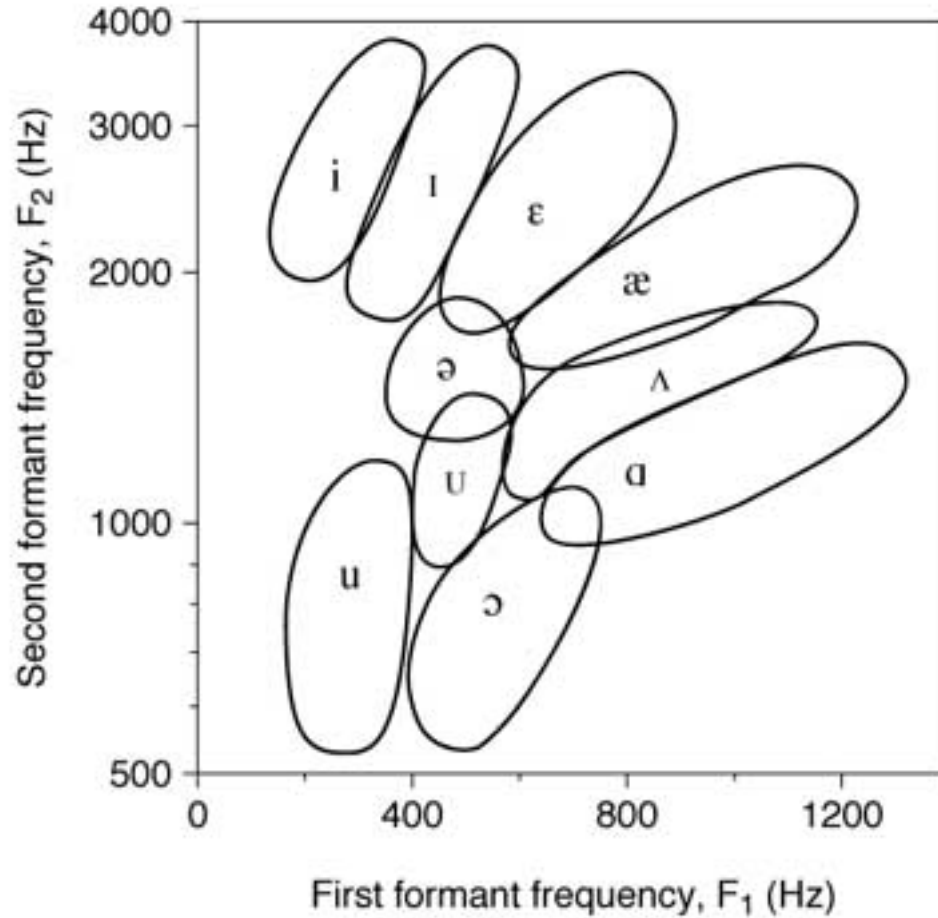
- Formants are defined as the spectra peaks of the sound spectrum
- Formants are independent of the F0 frequency, as they are defined over the envelope of the spectrum
- They are created by the pass of the sound through the vocal tract



Seeing formants: the spectrogram



Formants in the vowels



Example

