LING 520 Introduction to Phonetics I
Fall 2008

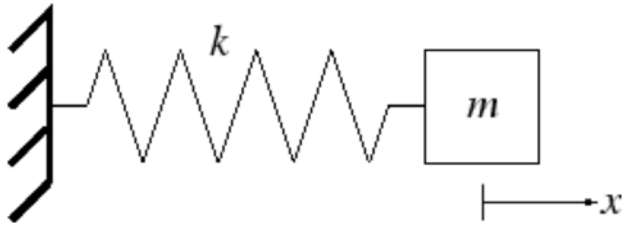
Week 4

Basic acoustics
Fourier transform and LTI systems
Recording and sampling

Oct. 1, 2008

Simple Harmonic Motion

- The ideal Mass-Spring System:



  - Hooke's law: F = −kx, where k is Spring Constant (a measure of the spring's stiffness) and x is Displacement.
  - System equation:

$$m\frac{d^2x}{dt^2} + kx = 0$$

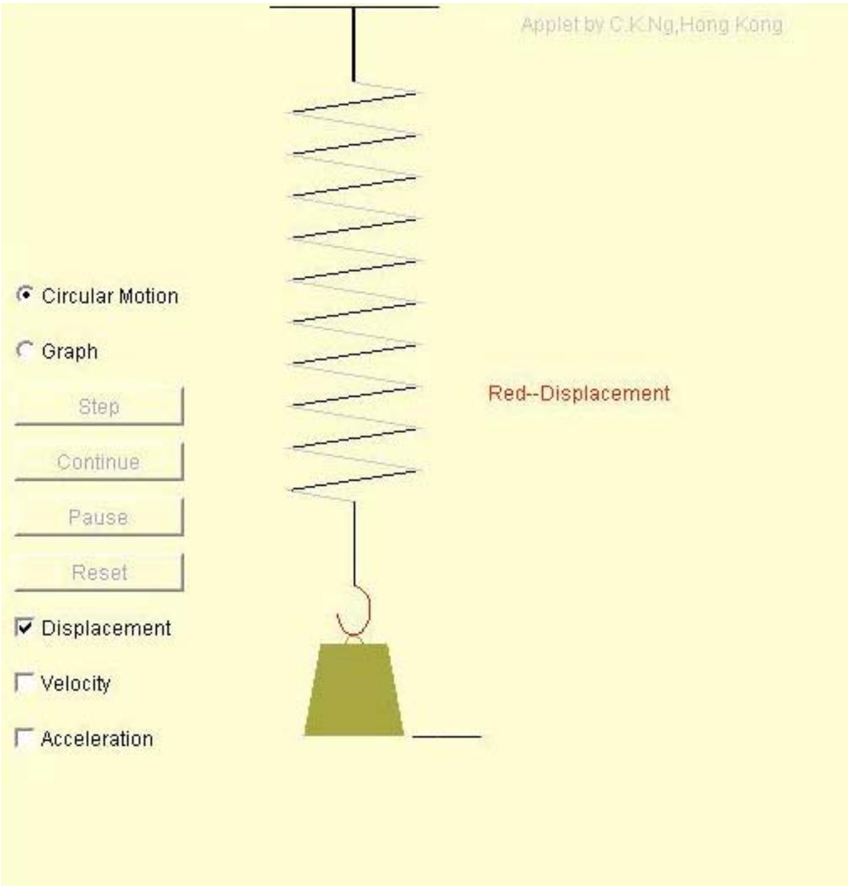
  - This second-order differential equation has solutions of the form:

$$x = A\cos(\omega_0 t + \phi), \omega_0 = \sqrt{k/m}$$

  - A and ϕ are determined by the initial displacement and velocity.
  - There are no losses in the system, so it will oscillate forever.

Simple Harmonic Motion

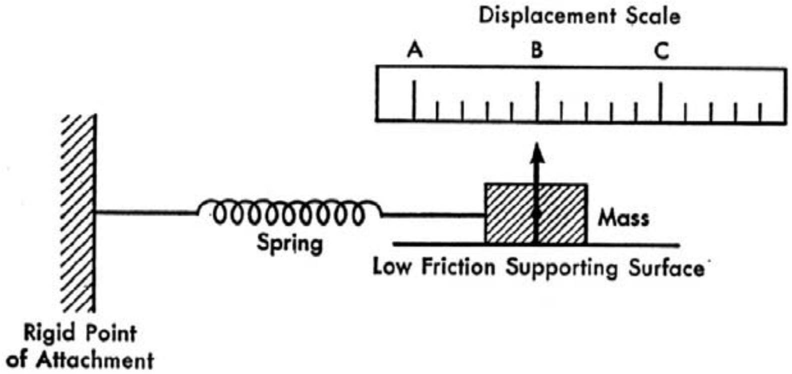
- **Demo:**



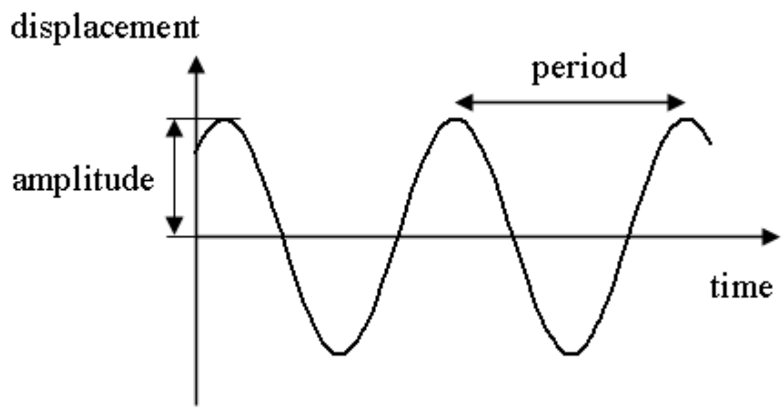
[from: C. K. Ng website]

Properties of vibrating systems

- *displacement*: momentary distance from restpoint B
- *cycle*: one complete oscillation
- *amplitude*: maximum displacement
- *frequency*: number of cycles per second (hertz or Hz)
- *period*: number of seconds per cycle
- *phase*: portion of a cycle through which a waveform has advanced relative to some arbitrary reference point

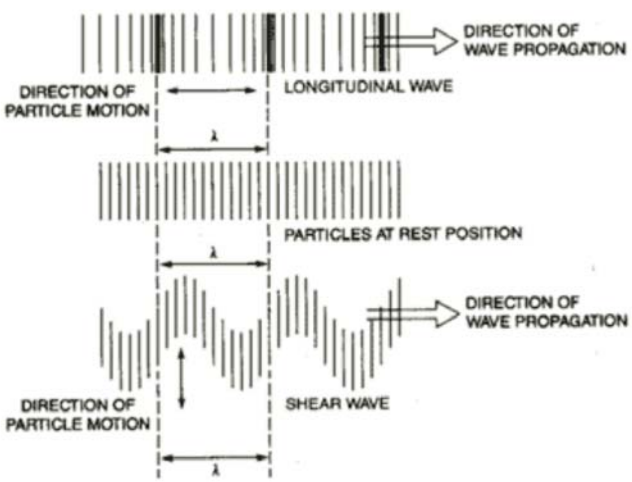


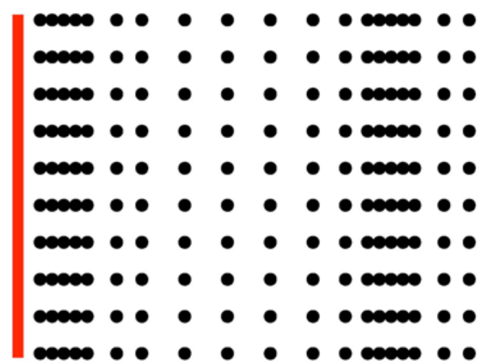
A simple spring-mass oscillator.

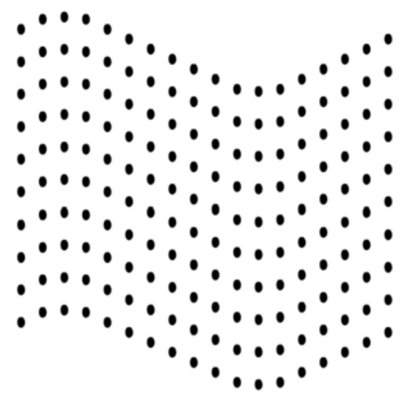


Waves

- A wave is an oscillation that spreads.
  The medium through which the wave travels
  experiences local oscillations as the wave passes,
  but the particles in the medium do not travel with the wave.
- Longitudinal Waves: Vibration of particles in the medium is parallel to the direction of wave propagation.
- Transverse Waves: Vibration of particles in the medium is perpendicular to the direction of wave.

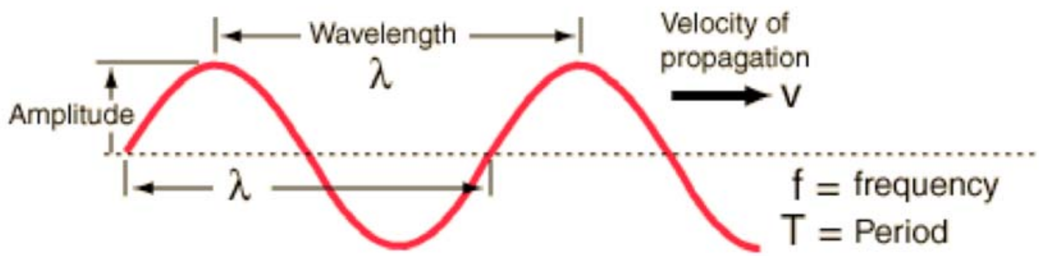






[Figure and animations from: NDT education website]

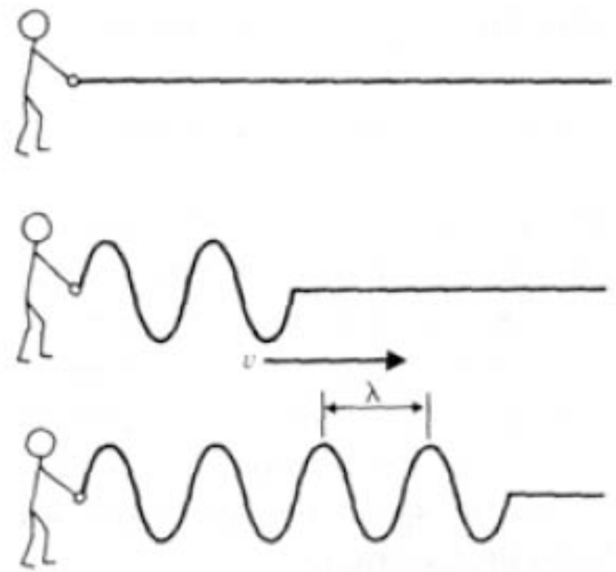
Waves



[Figure from: GSU physics website]

- Wavelength (λ): the distance between two successive points that are in phase, that is the points are doing the same thing at the same time (units: m).
- Frequency (f): frequency of the wave is the same as the frequency of its original source/oscillation.
- Velocity (V): propagation speed is dependent on the density and elasticity of the medium ($v_{solids} > v_{liquids} > v_{gases}$). At normal atmospheric pressure and a temperature of 20 degrees Celsius, a sound wave will travel at approximately 343 m/s.
- $V = f\,\lambda$

Waves



If I make f pulses per second, and the distance between each peak is λ (the "wavelength"), then in 1 sec, the wave has traveled a distance of $f\,\lambda$ so:

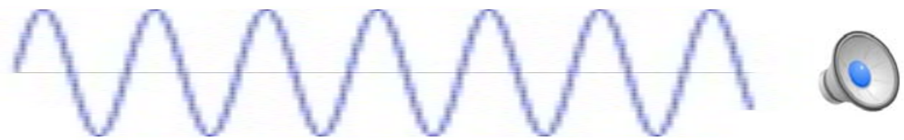
$$V = f\,\lambda$$

[from: Kathy Selby lecture notes]

Waves

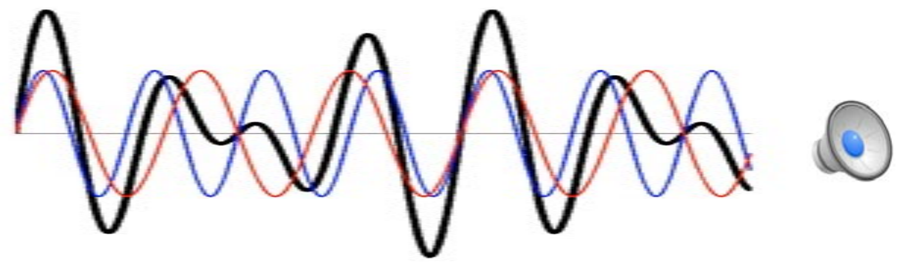
- Superposition: When two or more waves meet, they pass through each other. They are not changed afterwards. The actual displacement at any point is the sum of the individual displacements.

wave 1: 440 Hz



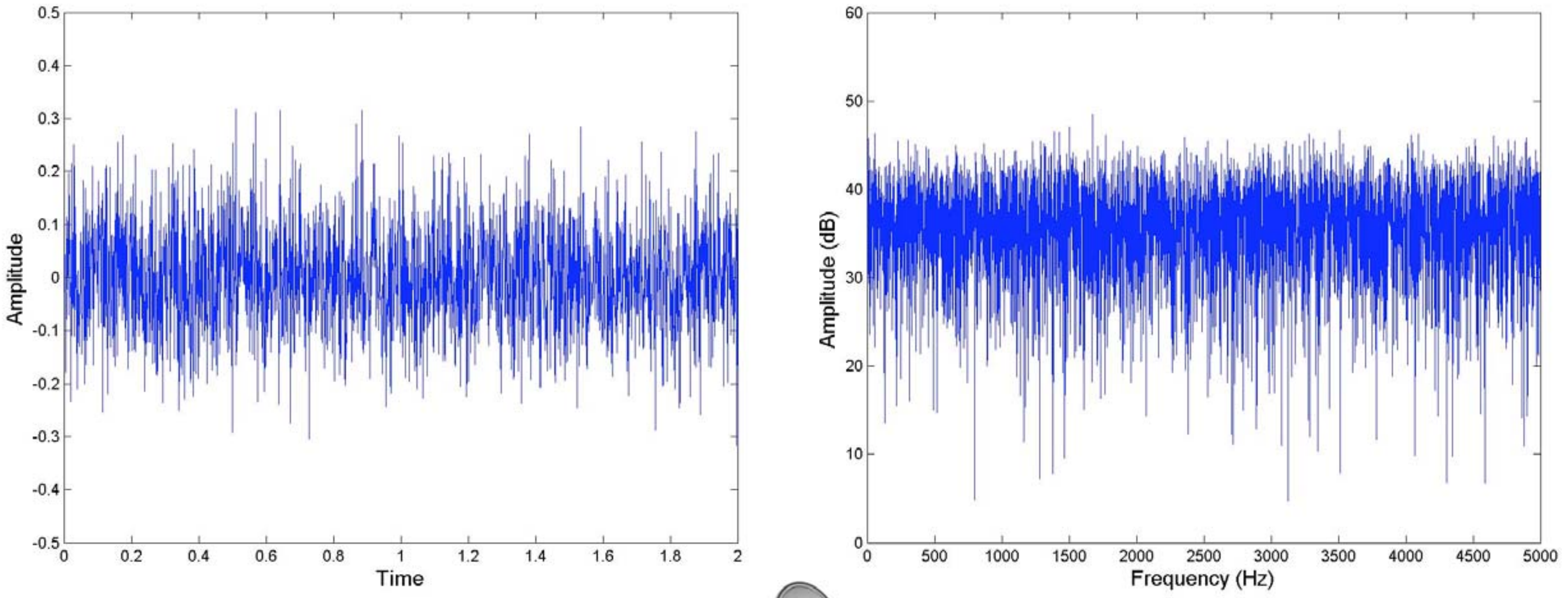
wave 2: 330 Hz

wave 1 + 2: ? Hz



Waves

- **Periodic vs. Aperiodic waves**
- **White noise**: a special signal that contains equal-amplitude components at all frequencies (as white light is composed of all of the colors of the spectrum of visible light), not just at integer multiples of a fundamental frequency. It is aperiodic.



Sound pressure and intensity

- "Threshold of audibility" (the minimum pressure fluctuation detected by the ear) is about 2 x 10^{-5}N/m^2 (Pascal) at 1000Hz; "Threshold of pain" correspond to a pressure 10^5 times greater than the audibility threshold. Because the ear is sensitive to such a wide range of pressure variations, sound pressure measurements are made on a logarithmic scale (decibel scale).
- Sound Pressure Level (SPL):

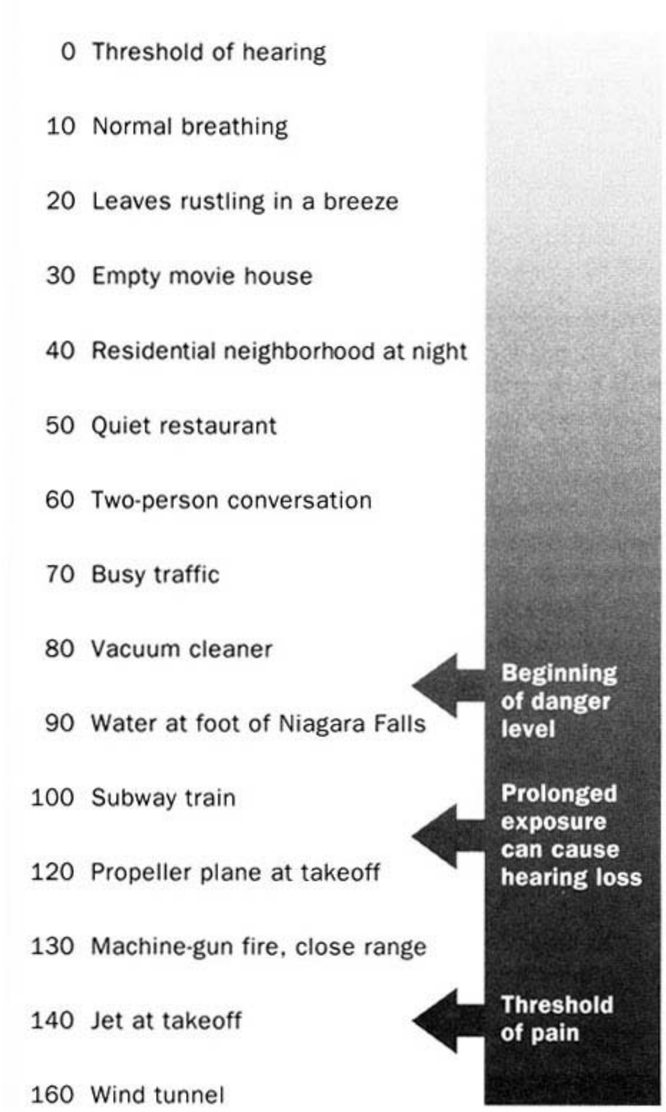
$$SPL = 20\log(p/p_0) = 10\log(p/p_0)^2,\ \ p_0 = 2\times 10^{-5}\ \mathrm{N/m}^2$$

- Sound Intensity Level (IL): The amount of sound energy which is transported past a unit area per unit of time (power per unit area).

$$IL = 10(\log I/I_0), I_0 = 10^{-12}\ \mathrm{watts/meter}^2$$

- Intensity is proportional to the square of pressure, therefore, SPL and IL are the same kind of unit. In a free field (no reflections), IL and SPL are nearly equal for a single source.

Decibel scale



Fourier transform

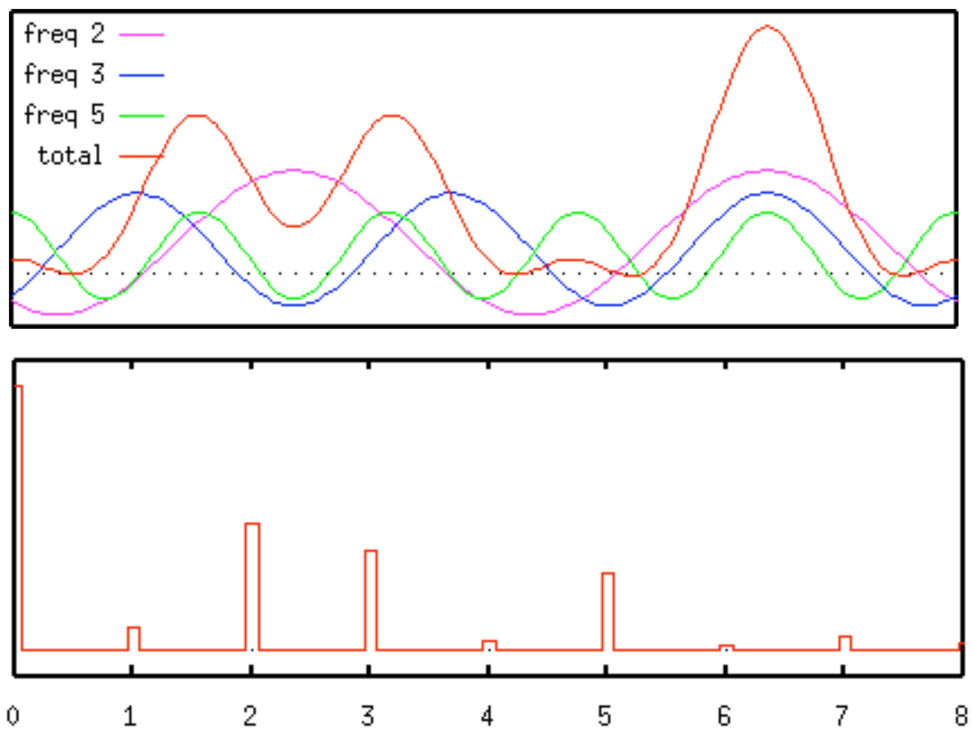
- **Fourier's theorem:**
  Any periodic signal (must satisfy Dirichlet conditions) is composed of a superposition of pure sine waves, whose frequencies are harmonics of the fundamental frequency of the signal.

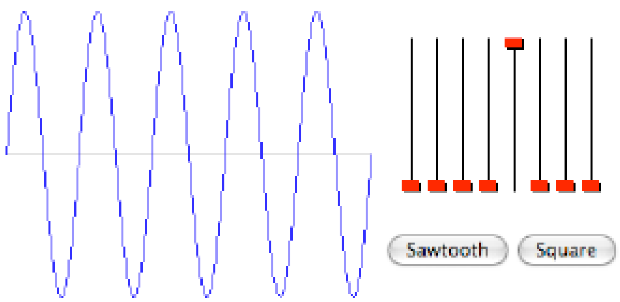
$$
\begin{aligned}
f(x) &= a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \\
&= a_0 + a_1 \cos x + a_2 \cos 2x + a_3 \cos 3x + \ldots \\
&\quad + b_1 \sin x + b_2 \sin 2x + b_3 \sin 3x + \ldots
\end{aligned}
$$

- The fundamental frequency F_0 determines the perceived pitch of the sound (In reality, perceived pitch is determined by the spacing harmonics as much as by F_0)
- The harmonic frequencies change the quality or timbre of the sound.

Spectrum

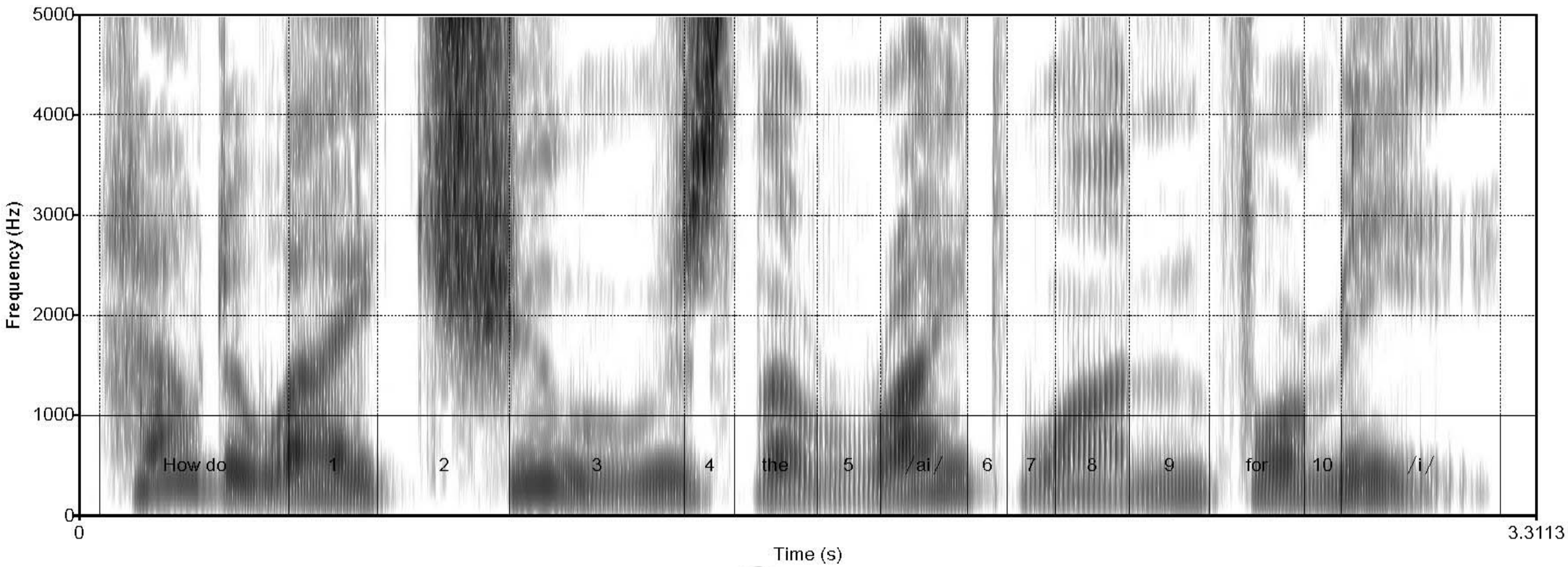
- A waveform can be represented as a spectrum showing:
  - Transmission of Component frequencies on the x-axis
  - Intensity (Amplitude) of each component on on the y-axis





Spectrogram

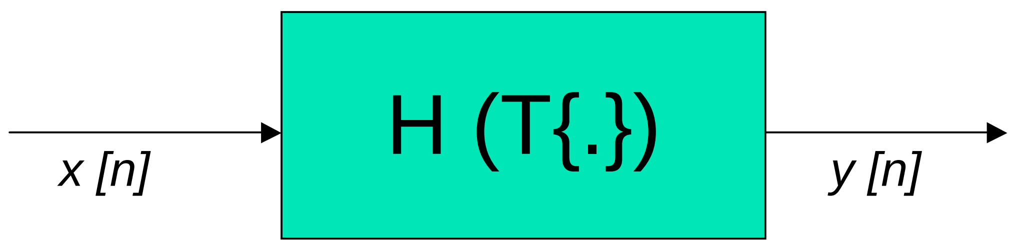
- A spectrogram displays:
  - Time along the x-axis
  - Component frequencies along the y-axis
  - Intensity (Amplitude) using darkness



Signals and Systems

- A signal is a real- or complex-valued function of time. Continuous time is represented by a real variable *t*, discrete time is represented by integer variable *n*.

- A single-input single-output (SISO) system is an "entity" that "processes" an input signal *x[n]* into an output signal *y[n]*.

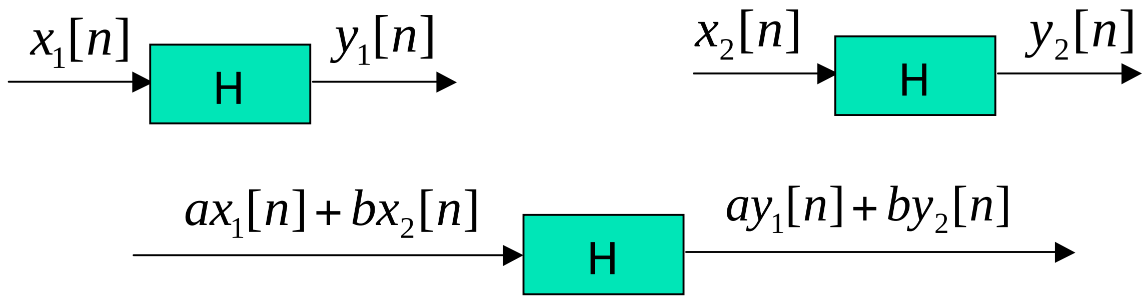


$$y[n] = T\{\ x[n]\ \}$$

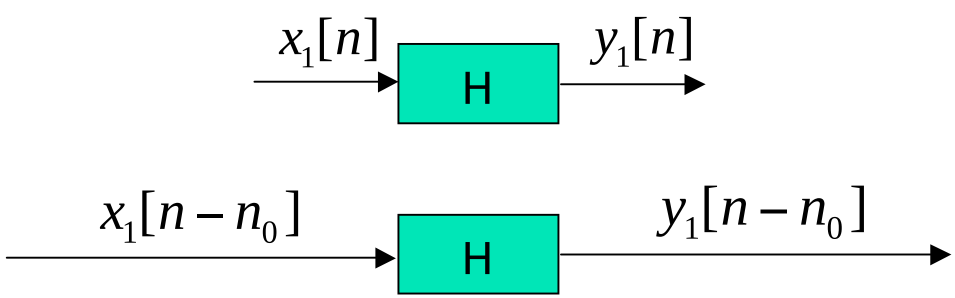
LTI Systems

- **Linear:**

$$T\{ax_1[n]+bx_2[n]\}=aT\{x_1[n]\}+bT\{x_2[n]\}$$



- **Time-Invariant:**



LTI Systems

- Discrete time unit impulse and impulse response:

$$\delta[n] = \begin{cases} 1 & \text{if } \; n = 0 \\ 0 & \text{otherwise} \end{cases} \qquad \delta[n] \longrightarrow \boxed{H} \longrightarrow h[n]$$

- Any discrete time signal can be decomposed as the sum of delayed unit impulses that are scaled by the signal value.

$$x[n] = \sum_{k=-\infty}^{\infty} (x[k]\delta[n-k]) \longleftrightarrow x[n] = x[n] * \delta[n]$$

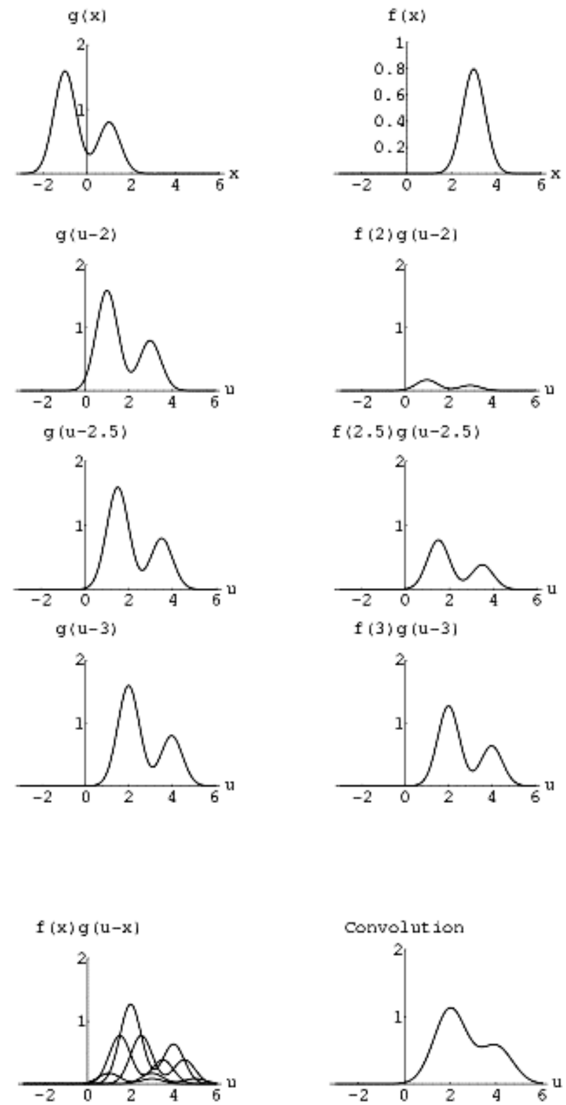
- LTI systems can be completely characterized by their impulse response (convolution), once we've found a system's impulse response, we know everyth9ong about the system's I/O behavior.

$$y[n] = T\left\{ \sum_{k=-\infty}^{\infty} x[k]\delta[n-k] \right\} \qquad y[n] = \sum_{k=-\infty}^{\infty} x[k]T\{\delta[n-k]\}$$

Convolution

$$C(\mathbf{u}) = f(\mathbf{x}) \otimes g(\mathbf{x}) = \int_{space} f(\mathbf{x})\, g(\mathbf{u}-\mathbf{x})\, d\mathbf{x}$$

$$= g(\mathbf{x}) \otimes f(\mathbf{x}) = \int_{space} g(\mathbf{x})\, f(\mathbf{u}-\mathbf{x})\, d\mathbf{x}$$



Convolution Theory

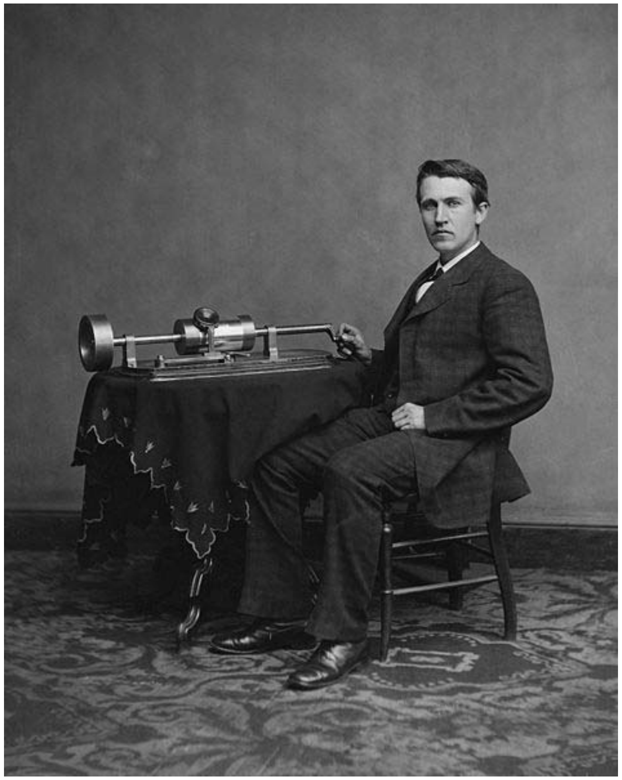
- The Fourier transform of a convolution is the product of the Fourier transforms
- The Fourier transform of a product is the convolution of the Fourier transforms

$$T(f \otimes g) = T(f)\,T(g)$$

$$T(f\,g) = T(f) \otimes T(g)$$

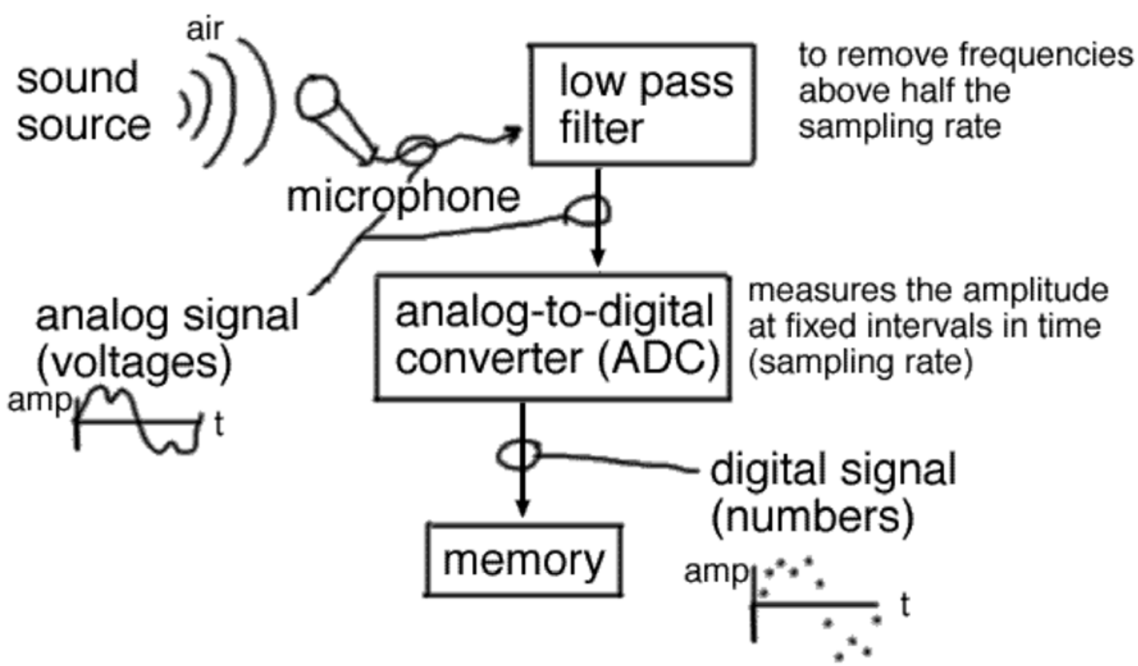
Recording

- **Phonograph:**



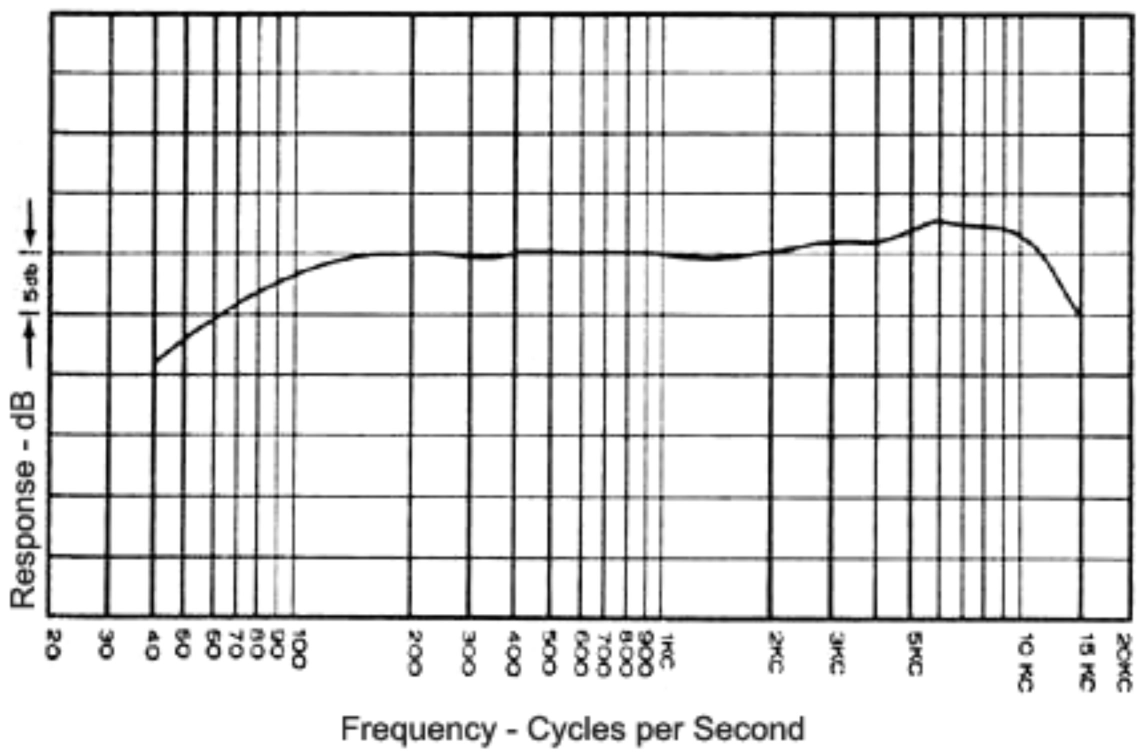
Recording

- **Digital recording:** The process of converting speech waves into computer-readable format is called *digitization*, or A/D conversion.



Recording

- **Microphone frequency response:** Frequency response refers to the way a microphone responds to different frequencies. An ideal "flat" frequency response means that the microphone is equally sensitive to all frequencies. In the real world a perfectly flat response is not possible and even the best "flat response" microphones have some deviation.



Recording

- **Signal to Noise Ratio (SNR):** Signal strength relative to background noise. The bigger the number, the better.

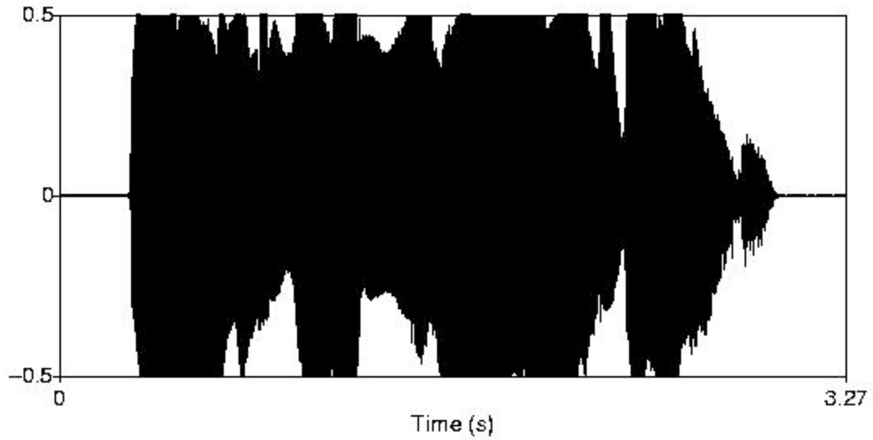
$$\mathrm{SNR(dB)} = 10\log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right) = 20\log_{10}\left(\frac{A_{\mathrm{signal}}}{A_{\mathrm{noise}}}\right)$$

- Classroom recording (SNR 29 dB)
- Laptop recording (SNR 44 dB)
- Professional recording (SNR 90 dB)

[From: Chilin Shih]

Recording

- **Clipping:** The sound is too loud (overloaded) for one or more components in the recording setup.



[From: Chilin Shih]

Sampling

- In order to transform sound into a digital format, you must sample the sound. The computer takes a snapshot of the sound level at small time intervals while you are recording.
- The number of samples taken each second is called the **sampling rate**. The more samples that are taken, the better sound quality. But we also need more storage space for higher quality sound.

44100 Hz

22050 Hz

11025 Hz

8000 Hz

5000 Hz

Sampling

- **Nyquist-Shannon theorem**:

  When sampling a signal (e.g., converting from an analog signal to digital), the sampling frequency must be greater than twice the highest frequency in the input signal in order to be able to reconstruct the original perfectly from the sampled version.

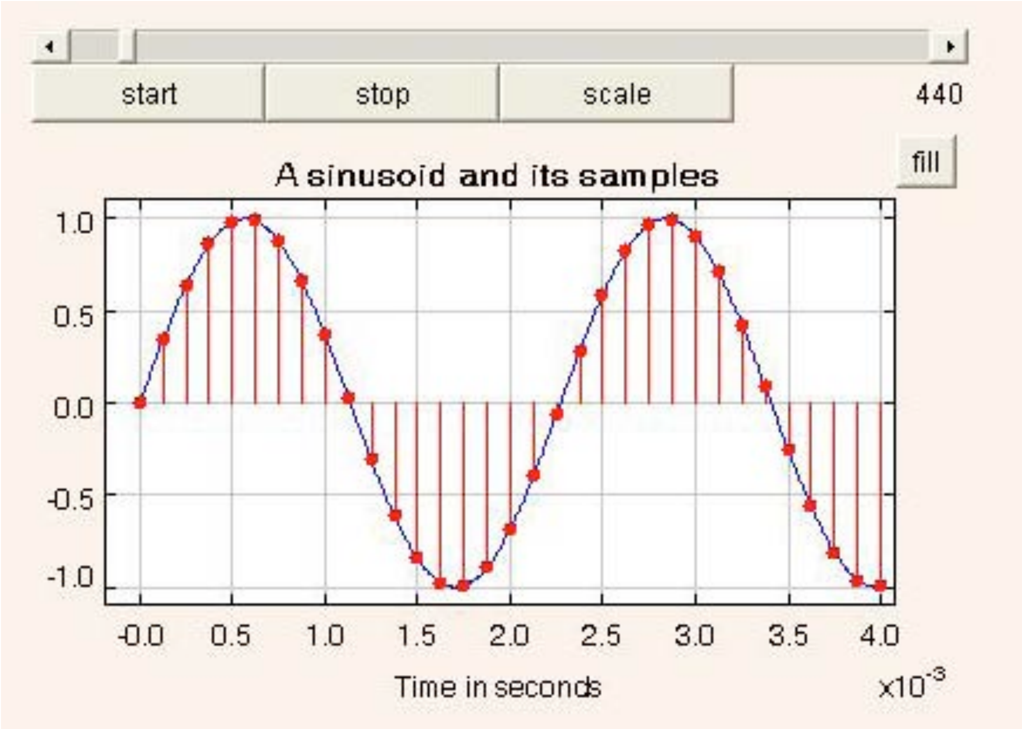
- **Nyquist frequency**: the highest frequency that can be coded at a given sampling rate in order to be able to fully reconstruct the signal, i.e., half of the sampling rate.

- **Aliasing**: If the sampling frequency is less than twice the highest frequency component, then frequencies in the original signal that are above half the sampling rate will be "aliased" and will appear in the resulting signal as lower frequencies.

- **Anti-Aliasing filter**: typically a low-pass filter that is applied before sampling to ensure that no components with frequencies greater than half the sample frequency remain.

Sampling

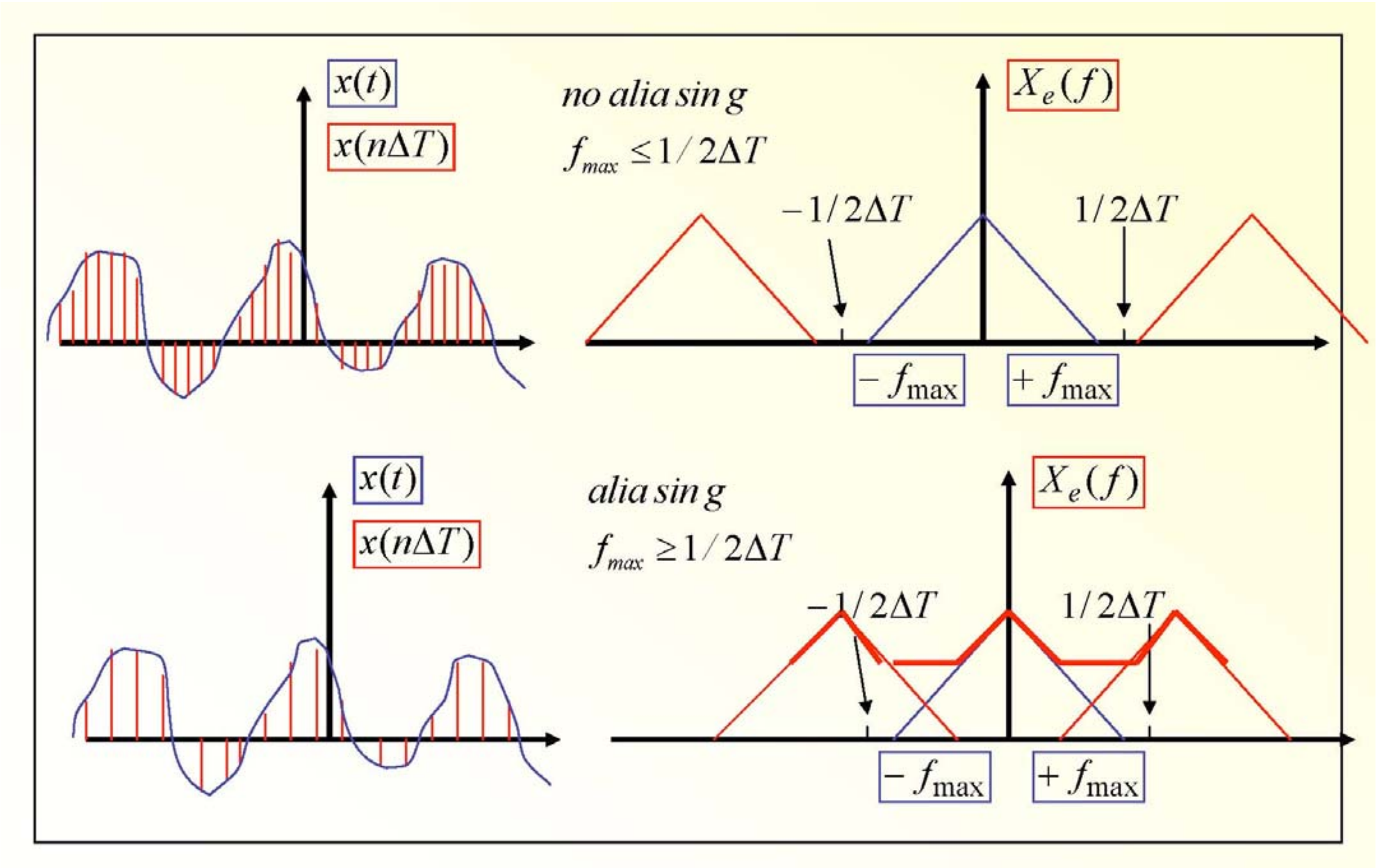
- Demo:



[from: Berkeley EECS website]

Sampling

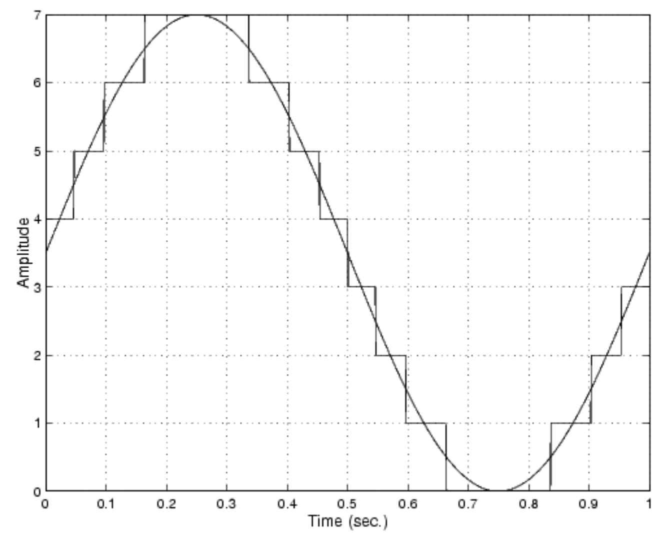
- Aliasing in frequency domain:



[From:http://meeng.technion.ac.il/Studies/PDF_ Files/5.Digitizing.pdf]

Quantization

- **Quantization:** Assigning a physical measurement to a binary number.
- **PCM** (Pulse-code modulation) is a digital representation of an analog signal where the magnitude of the signal is sampled at **uniform** intervals, then quantized to a series of binary numbers.
- If eight bits are allowed for the PCM sample, this gives a total of 256 possible values. PCM assigns these 256 possible values as 127 positive and 127 negative encoding levels, plus the zero-amplitude level (PCM assigns two samples to the zero level).



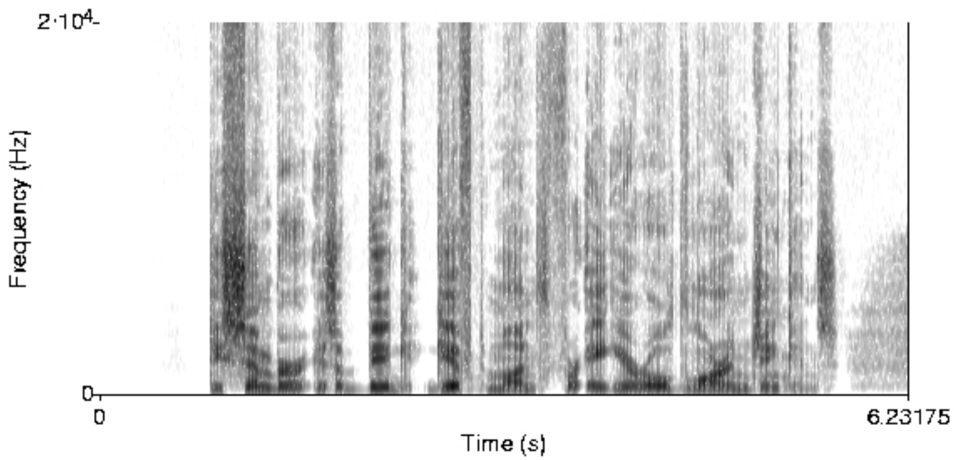
Sound file formats

Type	Extensions	Codec
AIFF (Mac)	.aif, .aiff	*PCM
AU (Sun/Next)	.au	*u-law
CD audio (CDDA)	N/A	PCM
Midi	.mid	NA
MP3	.mp3	MPEG Audio Layer-II
Windows Media Audio	.wma	Proprietary (Microsoft)
QuickTime	.qt, .mov	Proprietary (Apple Computer)
RealAudio	.ra, ram	Proprietary (Real Networks)
WAV	.wav	*PCM

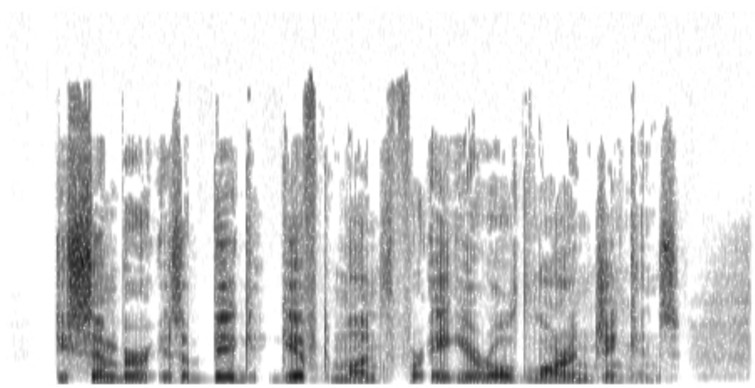
* Can be used with other codecs.

- The Wave file format stores information about the file's number of tracks (mono or stereo), sample rate, bit depth, as well as the uncompressed raw audio data (lossless).
- The advantages pf MP3 are the high compression rates (1/11 of the original size, still retaining considerable quality), the high availability of decoders and the low CPU requirements for playback (lossy).

Sound file formats



wav 550K



mp3 51K

wav->mp3->wav

[From: Chilin Shih]