CTP431- Music and Audio Computing
Fundamentals of Sound and Digital Audio

Graduate School of Culture Technology
KAIST
Juhan Nam

Outlines

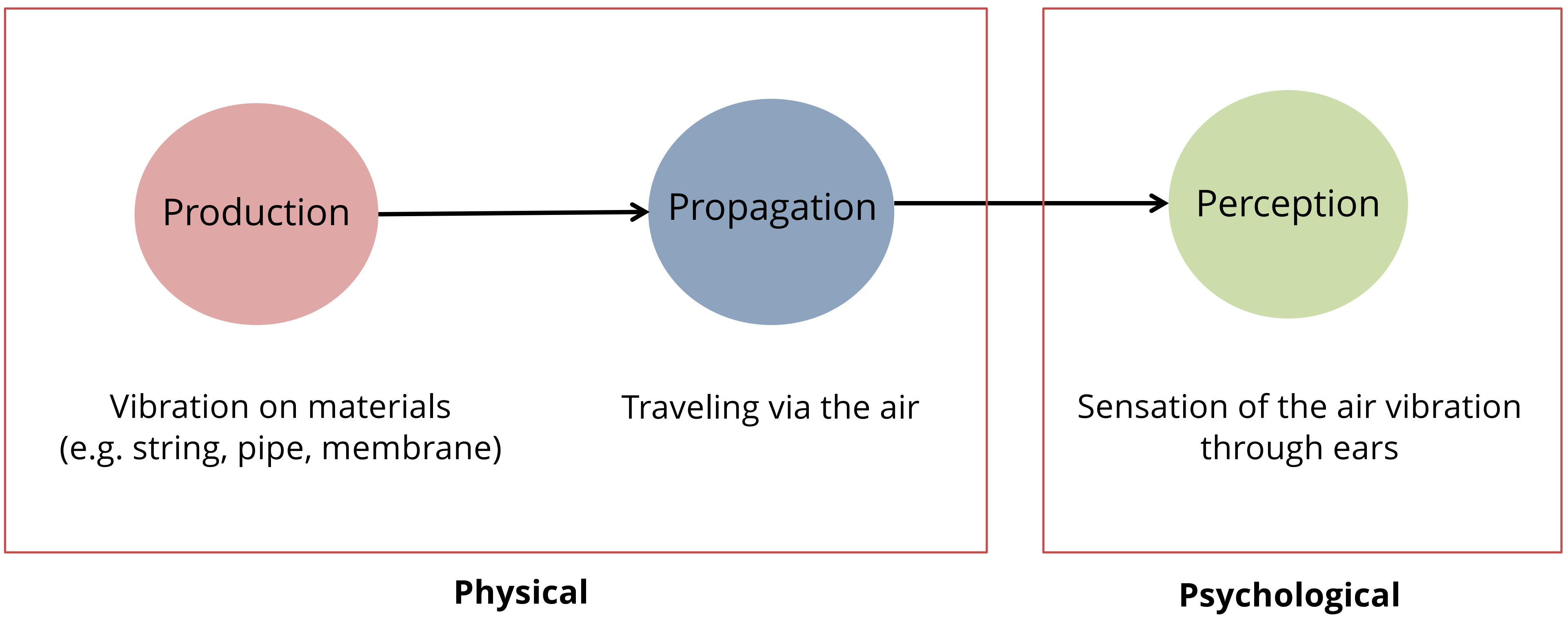
- What is Sound?

- Sound Properties
  - Loudness
  - Pitch
  - Timbre

- Digital Representation of Sound
  - Sampling
  - Quantization

What Is Sound?

- Vibration of air that you can hear
  - Compression and rarefaction of air pressure



Physical Sound

- Governed by “Newton’s law” and ”Wave” properties

- Sound production and propagation in musical instruments
  1. Drive force on a sound object
  2. Vibration by restoration force
  3. Propagation
  4. Reflection
  5. Superposition
  6. Standing Wave (modes): generate a tone
  7. Radiation from the object
  8. Propagation through air

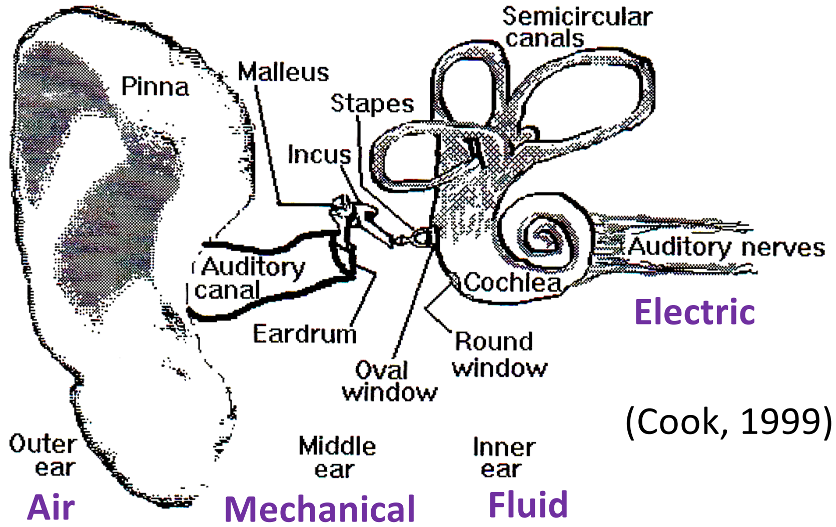
Demos

http://www.acs.psu.edu/drussell/demos.html

https://www.youtube.com/watch?v=_X72on6CSL0

Psychological Sound

- Governed by ears (physiological sense) and brain (cognitive sense)
  – human auditory system
- Ears
  – A series of highly sensitive transducers
  – Transform sound into subband signals
- Brain
  – Segregate and organize the auditory stimulus
  – Recognize loudness, pitch and timbre

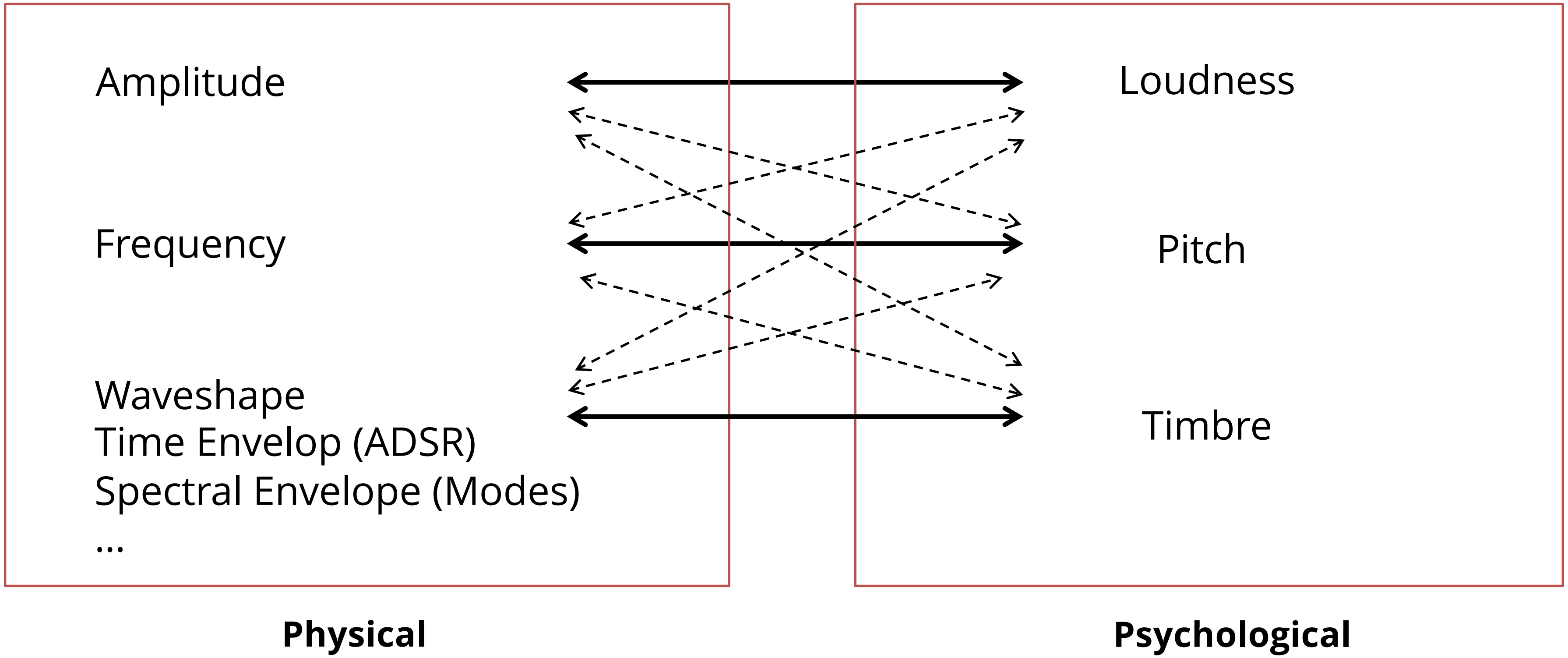


(Cook, 1999)

Auditory Transduction Video

http://www.youtube.com/watch?v=PeTriGTENoc

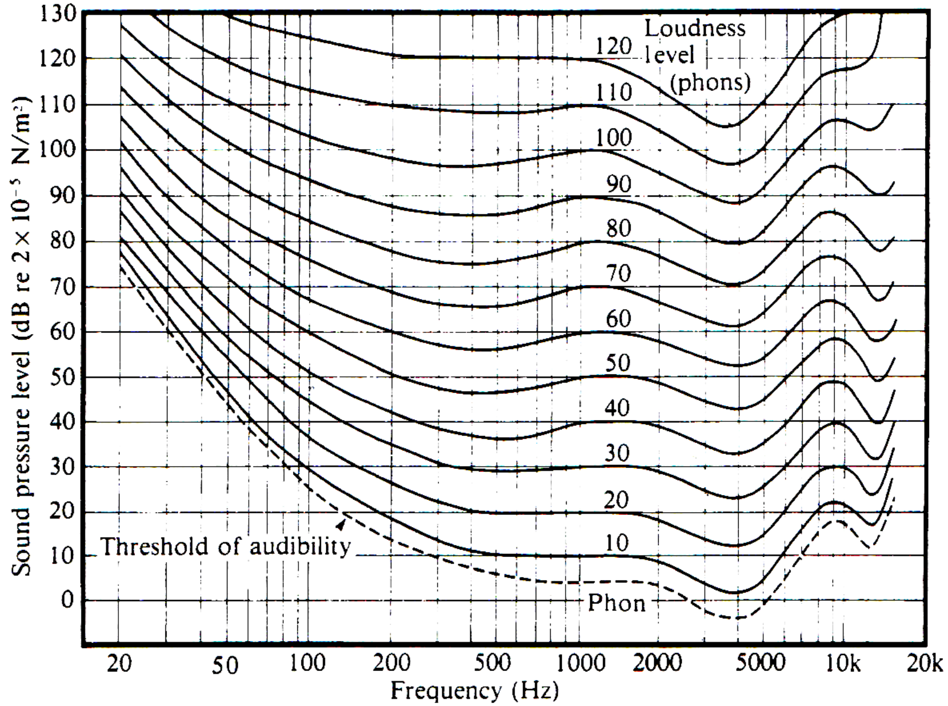
Sound Properties



Loudness

- Perceptual correlate of sound intensity
- Sound Pressure Level (SPL)
  - Objective measure of sound intensity
  - Log scale: $20\log_{10}(P/P_0)$

    $P_0 = 20\mu\text{Pa}$: threshold of human hearing
  - Loudness is proportional to SPL but not exactly
- Equal-Loudness Curve
  - Most sensitive to 2-5KHz tones
  - Threshold of hearing



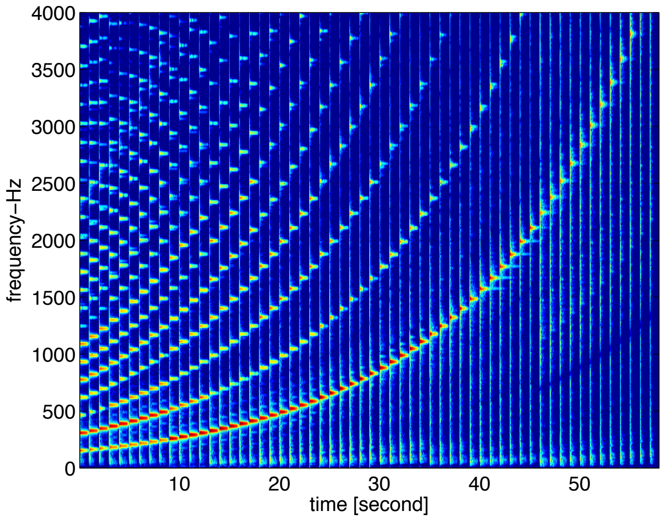
Equal-Loudness Curve
(also called Fetcher-Munson Curve)

Pitch

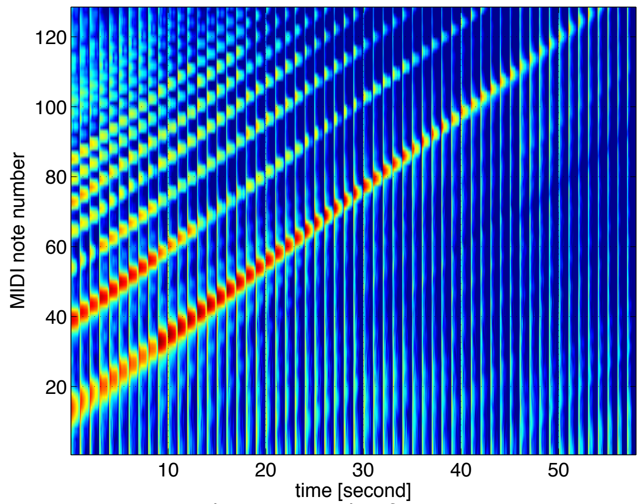
- Perceptual correlate of fundamental frequency (F0)

- Pitch Scale
  - Human ears are sensitive to frequency changes in a log scale
    - Ex) Piano note scale

- Frequency Range of Hearing
  - 20 to 20kHz



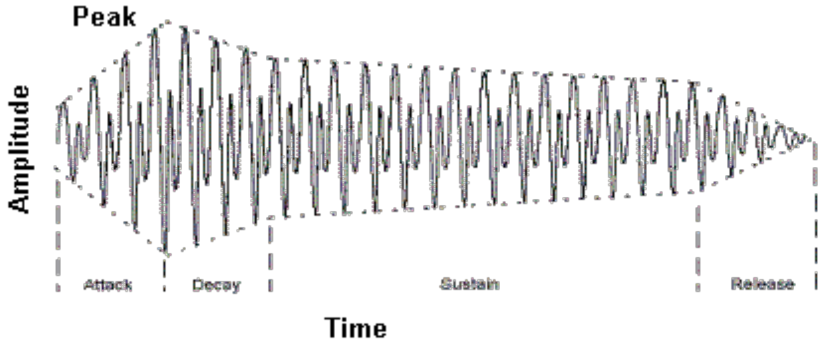
Chromatic Scale of Piano notes
(Linear Frequency)



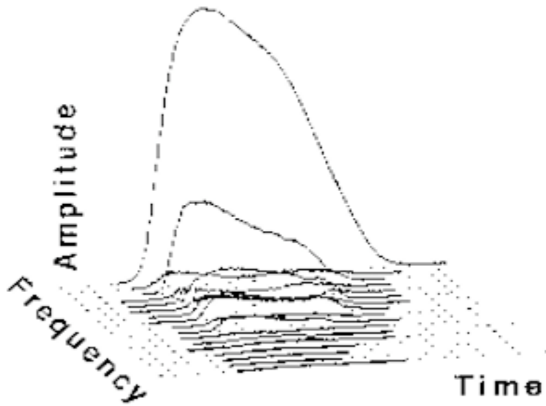
Chromatic Scale of Piano notes
(Log Frequency)

Timbre

- Related to identifying a particular sound object
  - Musical instruments, human voices, …
- Determined by multiple physical attributes
  - Time envelope (ADSR)
  - Spectral envelope
  - Changes of spectral envelope and fundamental frequency
  - Harmonicity: ratio between tonal and noise-like characteristics
  - The onset of a sound differing notably from the sustained vibration



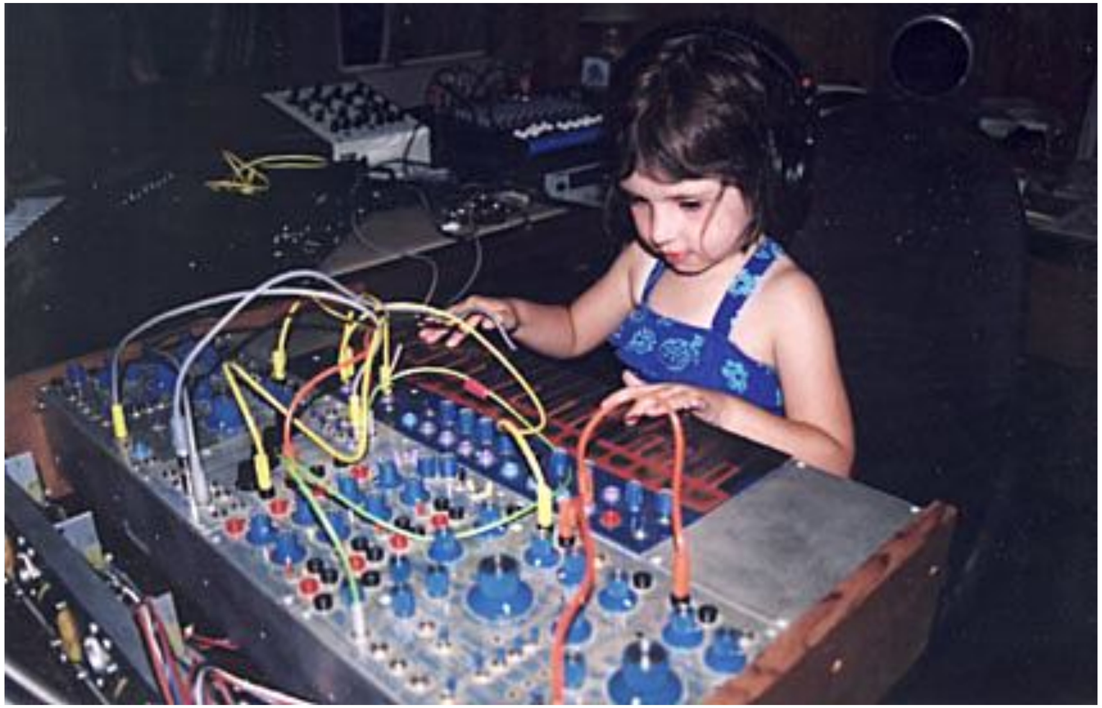
ADSR



Changes of spectral envelope

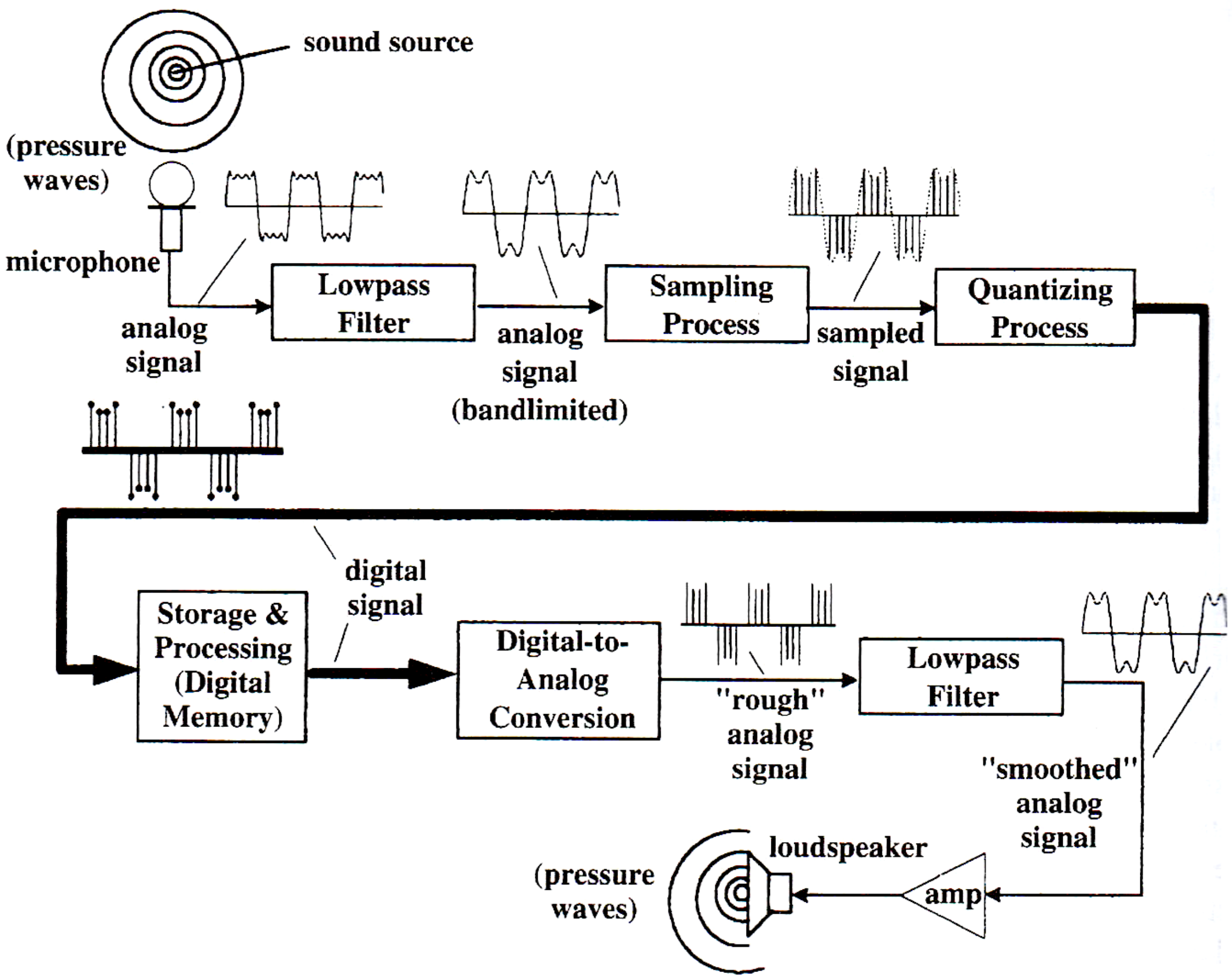
Timbre

- Determined by multiple parameters
  - Perspective of sound synthesis



Source: http://www.matrixsynth.com/2011/05/kid-with-buchla.html

Digital Audio Chain



Microphones / Speakers

- Microphones
  - Air vibration to electrical signal
  - Dynamic / condenser microphones
  - The signal is very weak: use of pre-amp

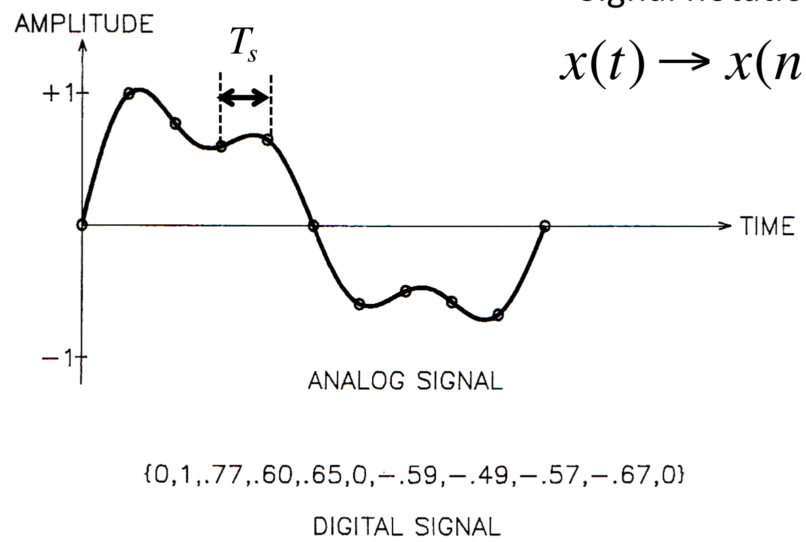
- Speakers
  - Electrical signal to air vibration
  - Generate some distortion (by diaphragm)
  - Crossover networks: woofer / tweeter

Sampling

- Convert continuous-time signal to discrete-time signal by periodically picking up the instantaneous values
  - Represented as a sequence of numbers; pulse code modulation (PCM)
  - Sampling period (T_s): the amount of time between samples
  - Sampling rate ($f_s = 1/T_s$)

Signal notation

$$x(t) \rightarrow x(nT_s)$$



Sampling Theorem

- What is an appropriate sampling rate?
  - Too high: increase data rate
  - Too low: become hard to reconstruct the original signal

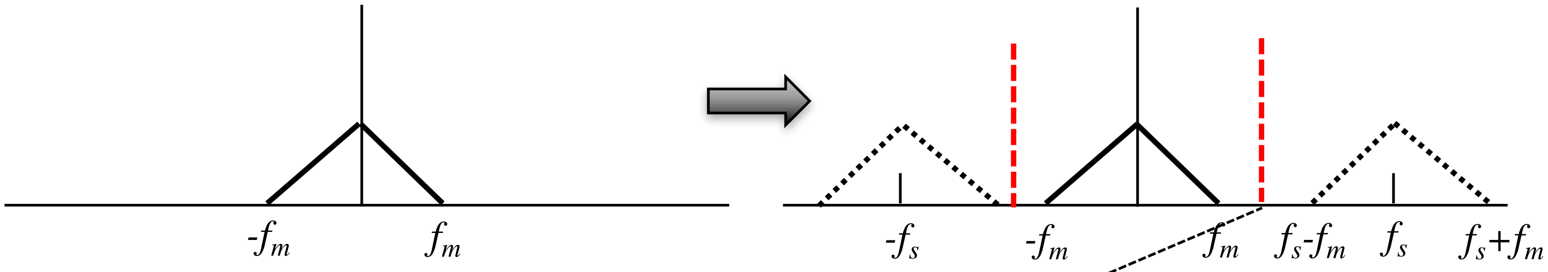
- Sampling Theorem
  - In order for a band-limited signal to be reconstructed fully, the sampling rate must be greater than twice the maximum frequency in the signal

$$\boxed{f_s > 2 \cdot f_m}$$

  - Half the sampling rate is called Nyquist frequency ($\frac{f_s}{2}$)

Sampling in Frequency Domain

- Sampling in time creates imaginary content of the original at every f_s frequency



To avoid overlap $\boxed{f_m < f_s - f_m}$

- Why ?

$$\boxed{f_2 = f_1 \pm m f_s}$$

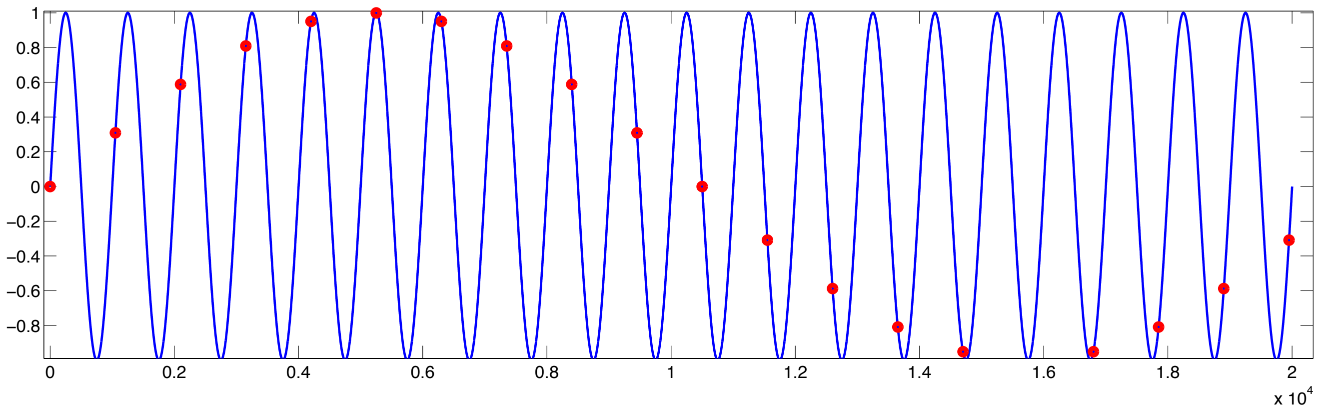
$$x_1(t) = A\sin(\omega_1 t) = A\sin(2\pi f_1 n / f_s)$$

$$x_2(t) = A\sin(\omega_2 t) = A\sin(2\pi f_2 n / f_s) = A\sin(2\pi (f_1 \pm m f_s) n / f_s)$$

$$= A\sin(2\pi f_1 n / f_s \pm 2\pi m n) = A\sin(2\pi f_1 n / f_s) = x_1(t)$$

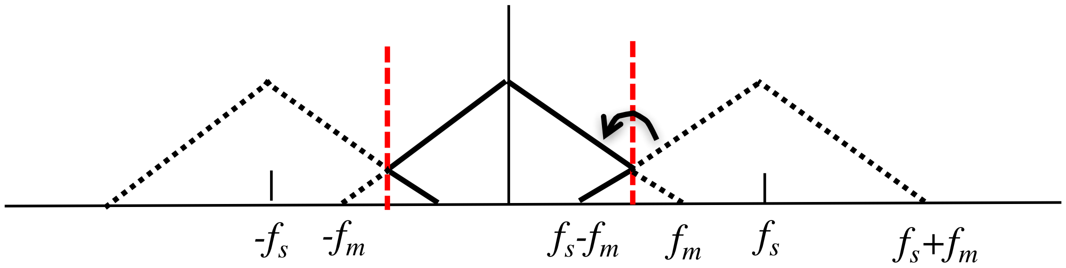
Aliasing

- If the sampling rate is less than twice the maximum frequency, the high-frequency content is folded over to lower frequency range

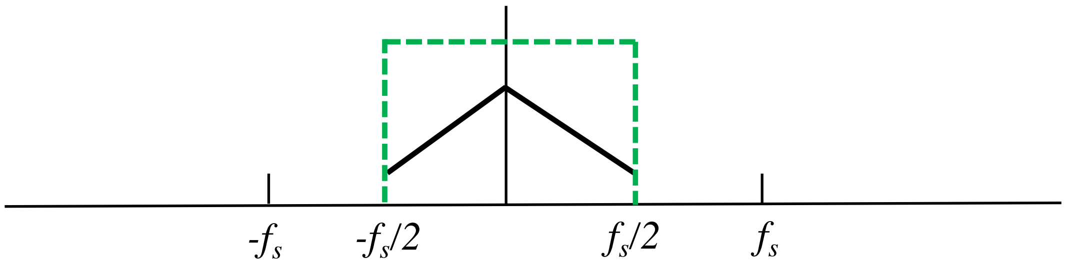


Aliasing in Frequency Domain

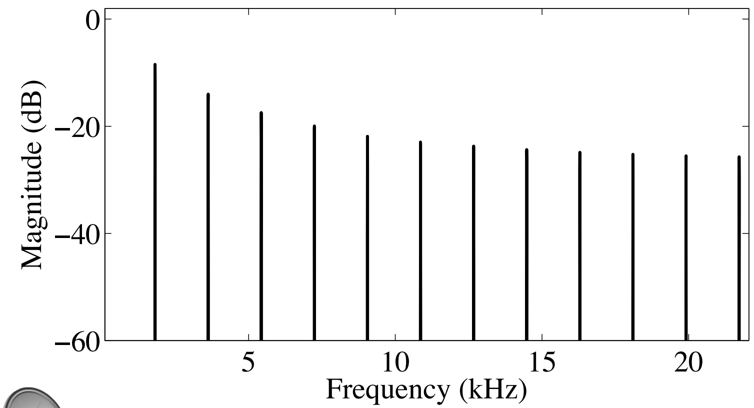
- The high-frequency content is folded over to lower frequency range from the replicated images



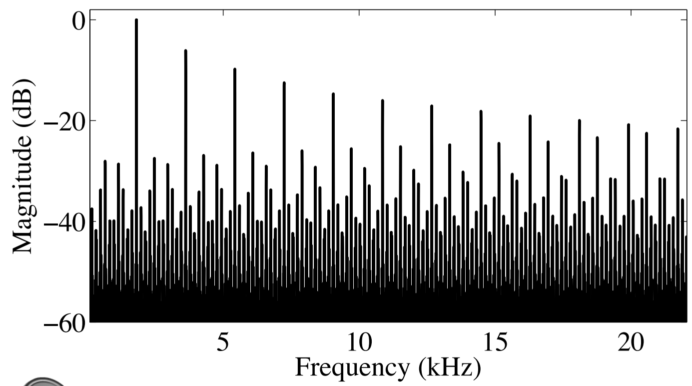
- A low-pass filter is applied before sampling to avoid the aliasing noise



Example of Aliasing

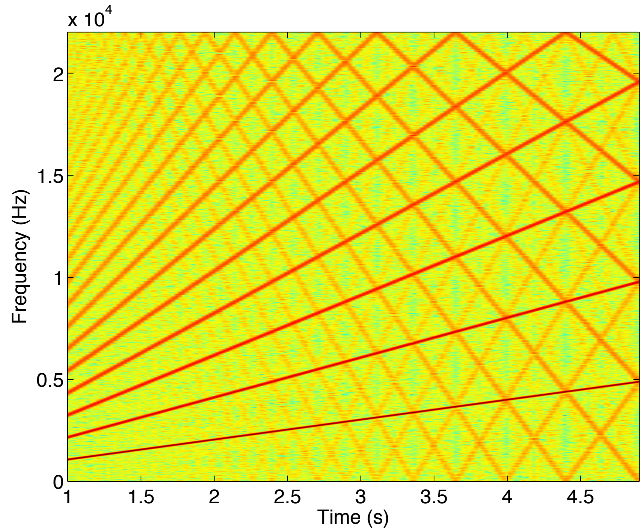


Bandlimited sawtooth wave spectrum



Trivial sawtooth wave spectrum

Frequency sweep of the trivial sawtooth wave



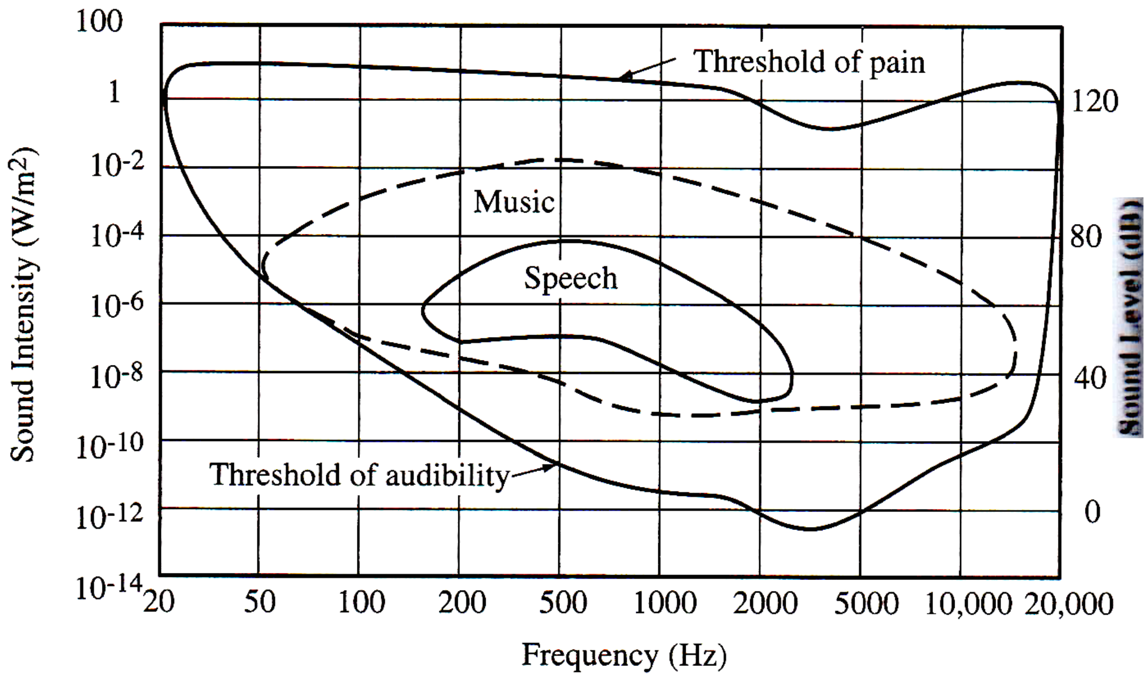
Example of Aliasing

- Aliasing in Video
  - https://www.youtube.com/watch?v=QOqtdl2sJk0
  - https://www.youtube.com/watch?v=jHS9JGkEOmA

(Note that video frame rate corresponds to the sampling rate)

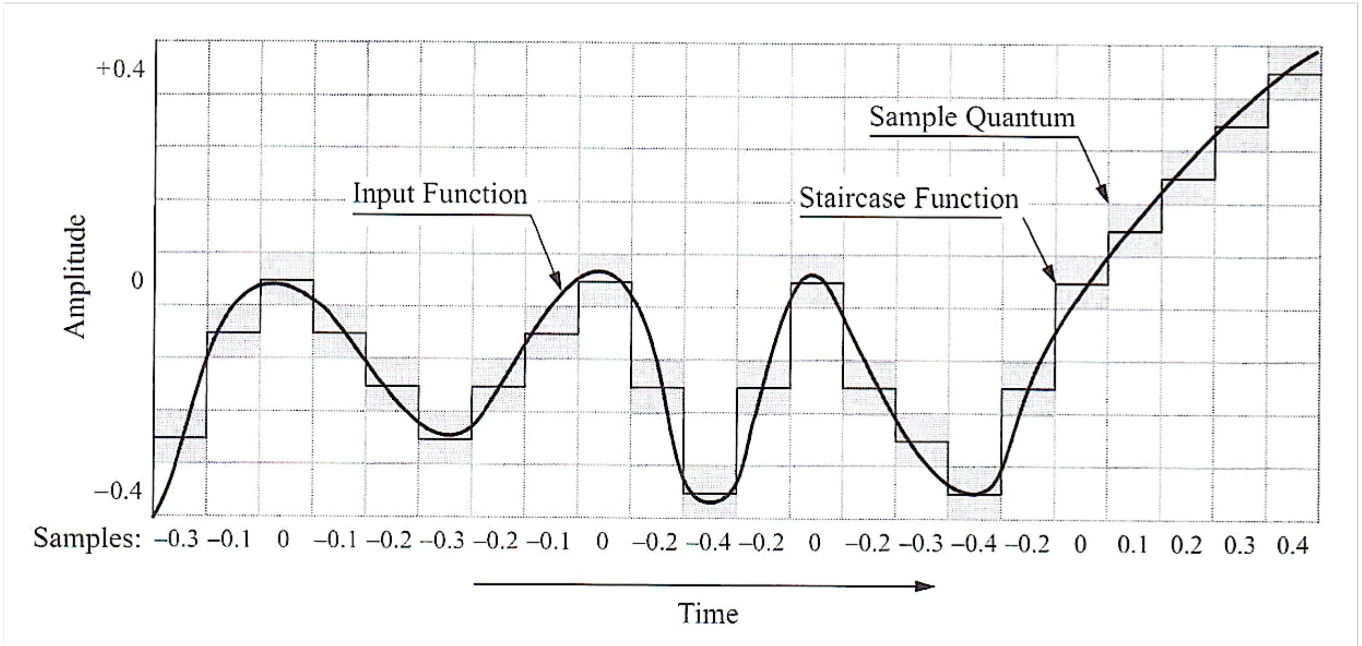
Sampling Rates

- Determined by the bandwidth of signals or hearing limits
  - Consumer audio product: 44.1 kHz (CD)
  - Professional audio gears: 48/96/192 kHz
  - Speech communication: 8/16 kHz



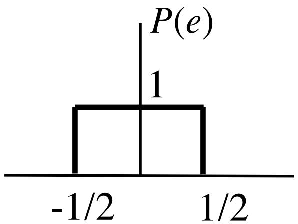
Quantization

- Discretizing the amplitude of real-valued signals
  - Round the amplitude to the nearest discrete steps
  - The discrete steps are determined by the number of bit bits
    - Audio CD: 16 bits (-2^{15} ~ $2^{15}-1$) ← B bits (-2^{B-1} ~ $2^{B-1}-1$)



Quantization Error

- Quantization causes noise
  - Average power of quantization noise: obtained from the probability density function (PDF) of the error



Root mean square (RMS) of noise

$$\sqrt{\int_{-1/2}^{1/2} x^2 p(e)\,dx} = \sqrt{\tfrac{1}{12}}$$

- Signal to Noise Ratio (SNR)
  - Based on average power

RMS of full-scale sine wave

$$20\log_{10}\frac{S_{\mathrm{rms}}}{N_{\mathrm{rms}}} = 20\log_{10}\frac{2^{B-1}/\sqrt{2}}{\sqrt{\tfrac{1}{12}}} = 6.02B + 1.76\ \mathrm{dB}$$ (With 16bits, SNR = 98.08dB)

  - Based on the max levels

$$20\log_{10}\frac{S_{\max}}{N_{\max}} = 20\log_{10}\frac{2^{B-1}}{\tfrac{1}{2}} = 6.02B\ \mathrm{dB}$$ (With 16bits, SNR = 96.32 dB)