Speech Processing (Part I)

CSCI-B 659 14328

THURSDAY, AUGUST 29, 2019

LECTURE #2

Learning Objectives

- **You the student will be able to:**
  - Explain how sound is produced from air particles
  - Understand how microphones generate electrical signals from sound particles
  - Describe audio sampling
  - Describe audio quantization

Video on Sound Production

Capturing Sound: Microphones

"Transducer that converts air pressure variations into an electrical signal"

All microphones have diaphragms

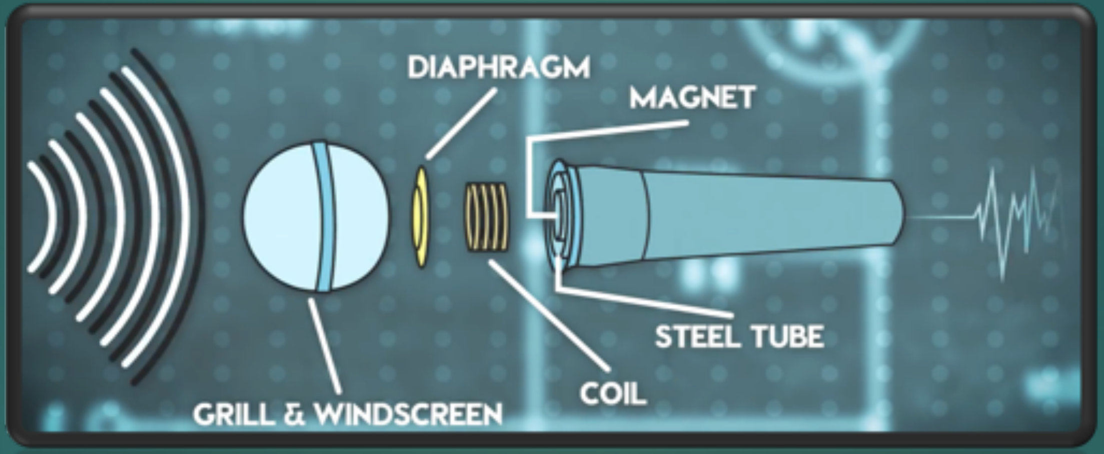
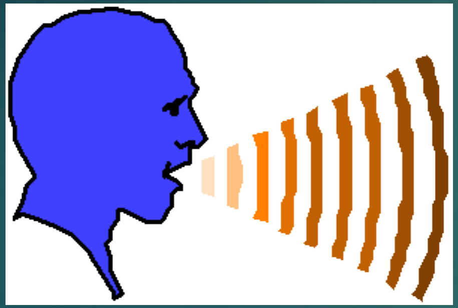


Image courtesy of LinkedIn Learning Solutions

- Two main types of microphones:
  - ***Dynamic***: Coil-based microphones
    - All-purpose, often used during on-stage performances
    - Sturdy, less expensive, capture less detail (less sensitive)
    - Useful for recording instruments
  - ***Condenser***: Capacitor based
    - Useful for vocal recording in recording studios
    - Superior sound quality, more sensitive, more expensive
    - Also used in cell phones and computers

Capturing Sound: Microphones



Speaker

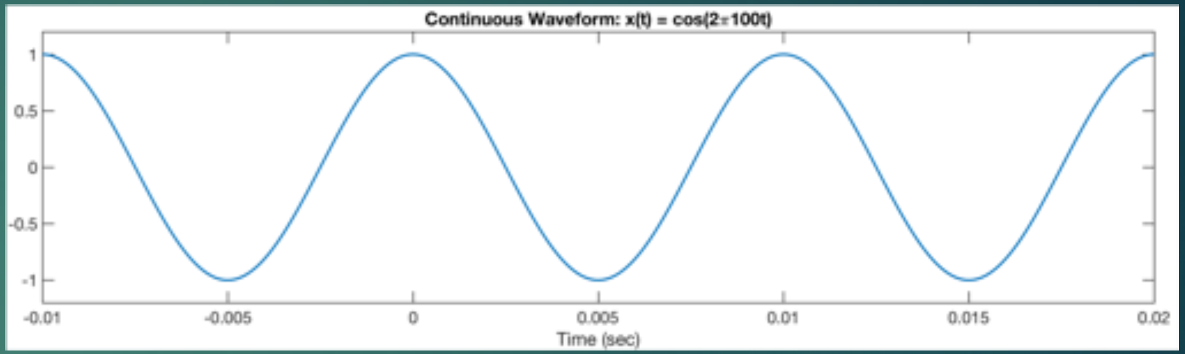
Why use an array of microphones?

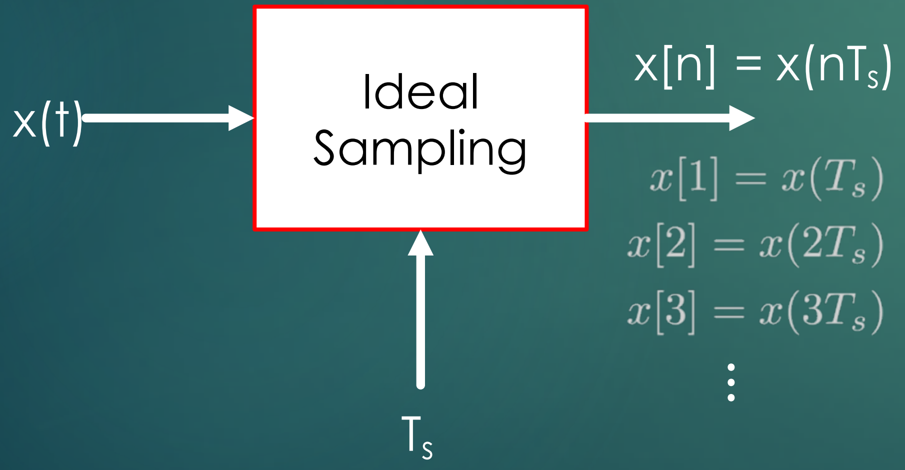
Microphone

- Microphones output a electrical (analog) signal, $x(t)$
  - $x(t)$ has values at all time t (i.e. at 2, 0.09, or 0.00003 seconds)
- Computers and most devices require digital signals
- We have to ***sample*** sound before it can be used by computers.

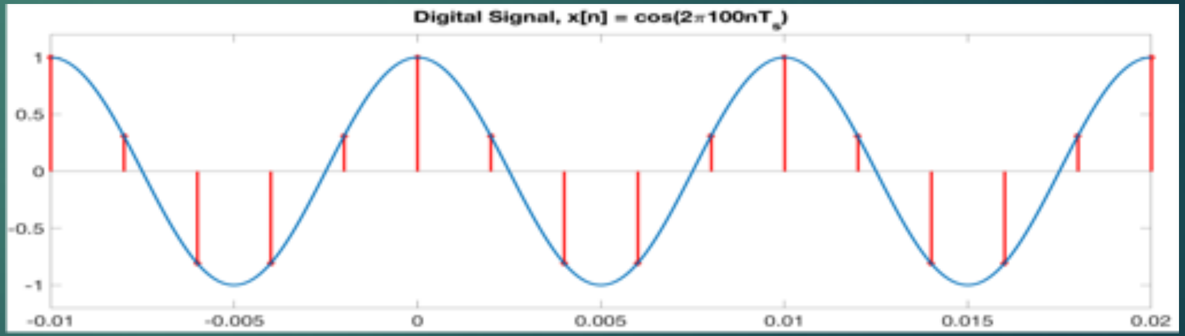
Audio Sampling

- Sampling converts an analog sound wave into a sequence of discrete sample values
  - $x(t)$ – analog signal
  - $x[n]$ – discrete signal





$$x[1] = x(T_s)$$
$$x[2] = x(2T_s)$$
$$x[3] = x(3T_s)$$
$$\vdots$$

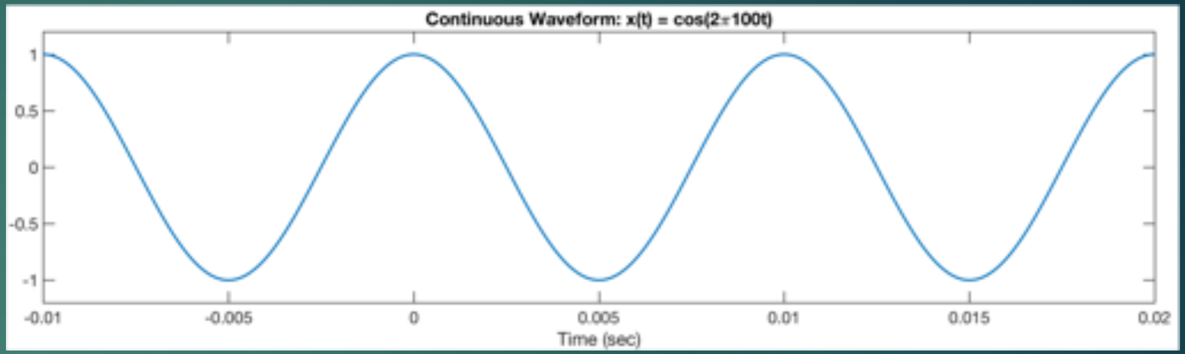


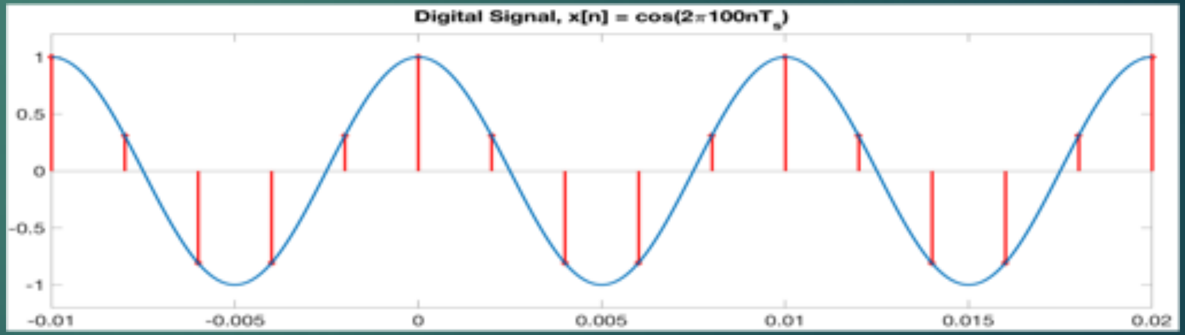
T_s: sampling period (seconds)

Think of this as a switch that turns on every T_s seconds

Audio Sampling (cont.)

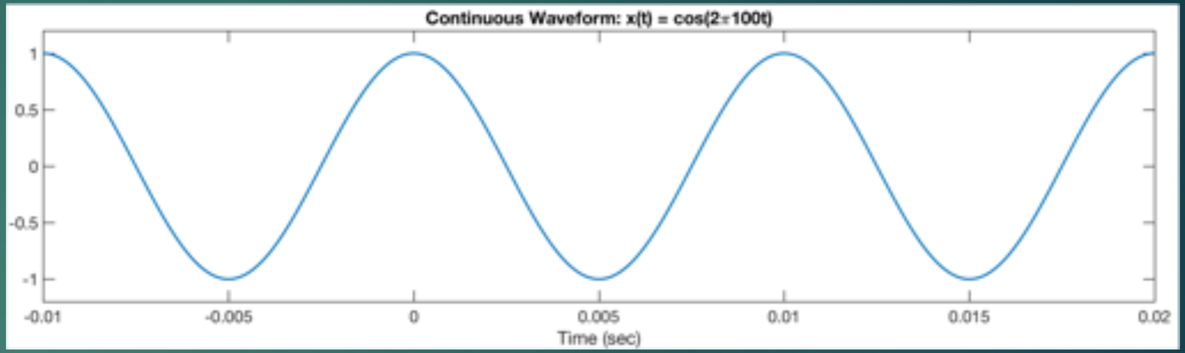
- T_S: sampling period (seconds)
- Sampling frequency (or rate)
  - $F_s = 1/T_s$
  - Unit is Hertz (Hz): samples per second
- If my sampling period is .01 seconds, how many discrete samples are there in a 3.5 second signal?
  - What is it for a .5 second sampling period?

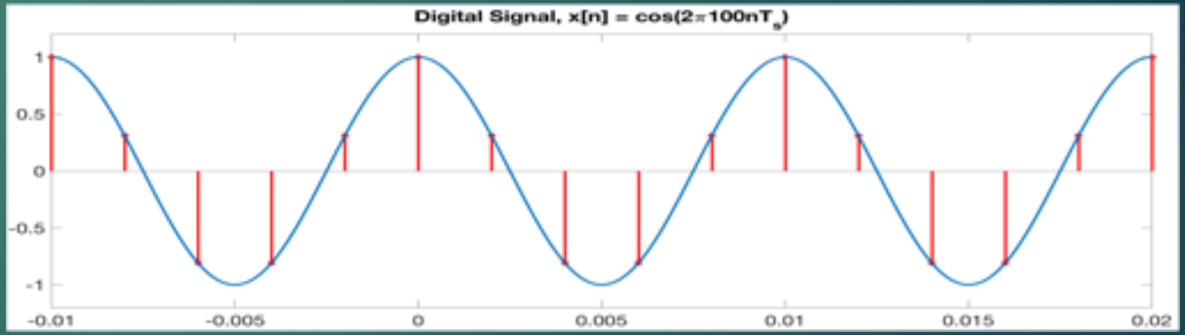




Audio Sampling (cont.)

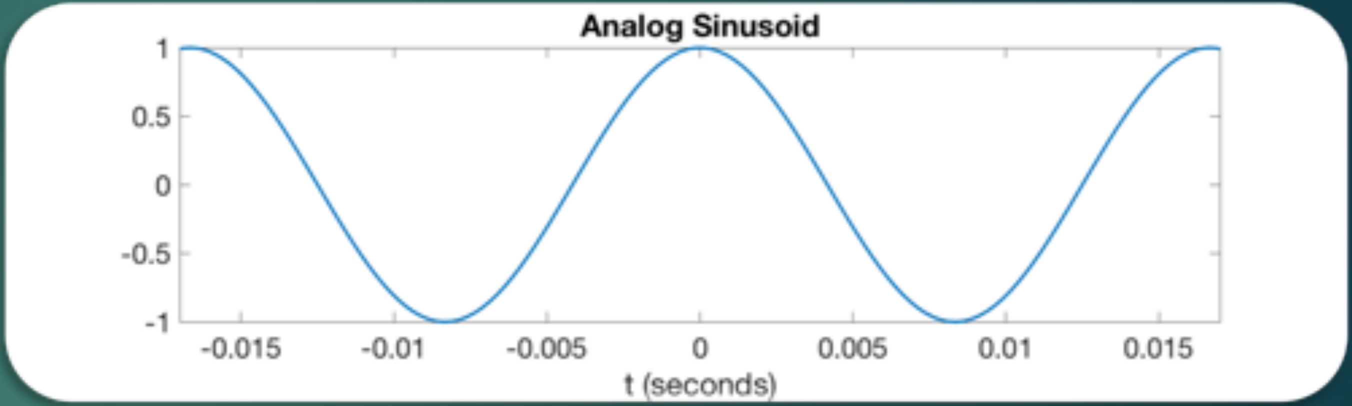
- Goal:
  - Sample the audio signal at an *appropriate* sampling rate
  - This produces a "good" representation of the continuous signal
- Unfortunately, if we sample too slow (e.g. T_s is too large or F_s is too small) this does not happen!
  - This is called ***aliasing***.



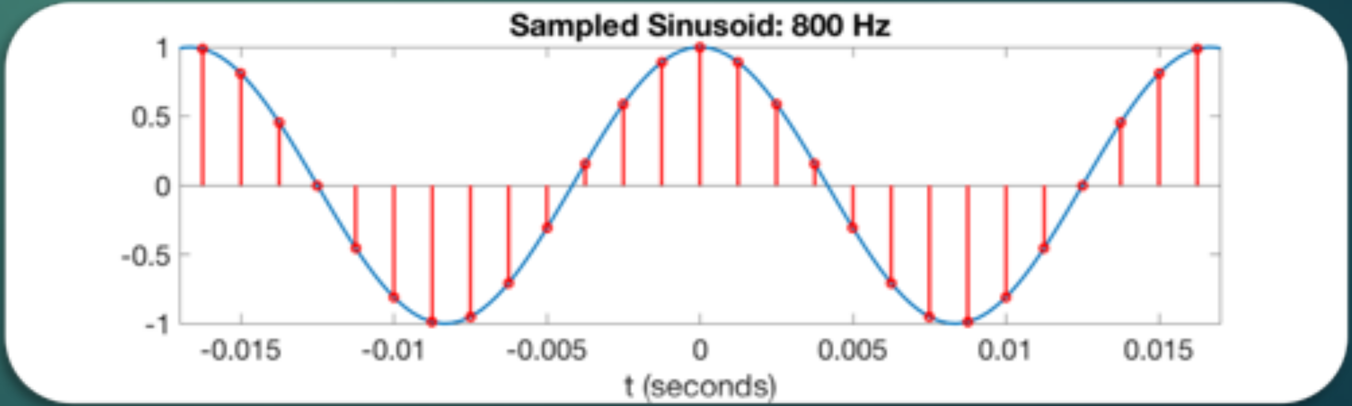


Example: Effects of Aliasing

- Analog Sinusoid
  - x(t) = cos(2πf*t)
  - f = 60 Hz



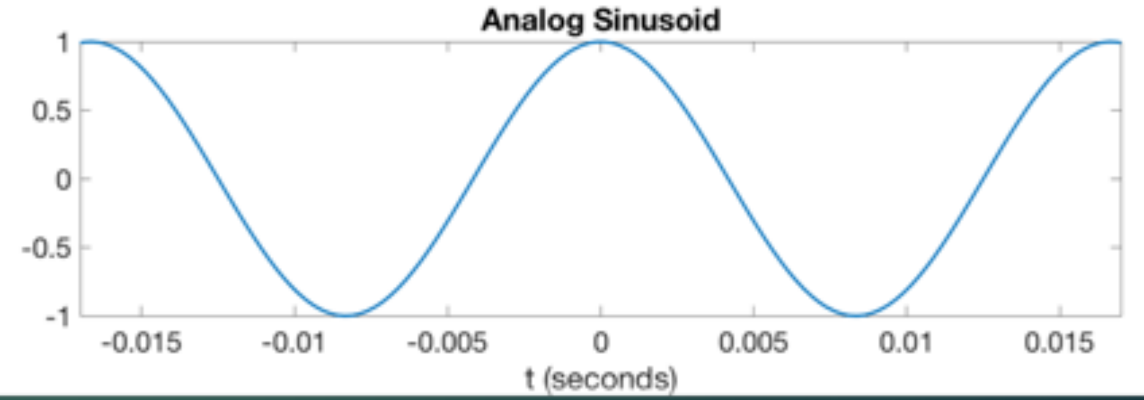
- Let's sample the sinusoid
  - F_S = 800 Hz (Ts = 0.0013s)



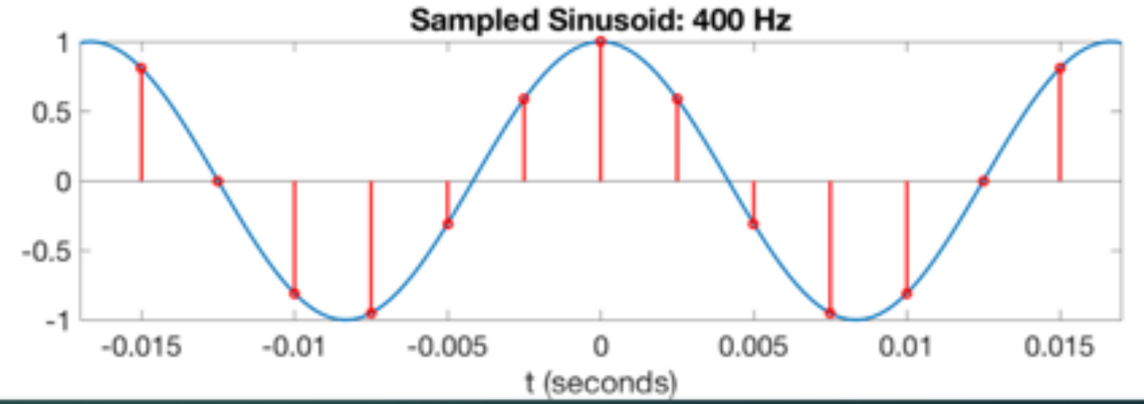
The outline of the sampled signal matches the analog signal

Example: Effects of Aliasing

- Analog Sinusoid
  - $x(t) = \cos(2\pi f^*t)$
  - $f = 60$ Hz



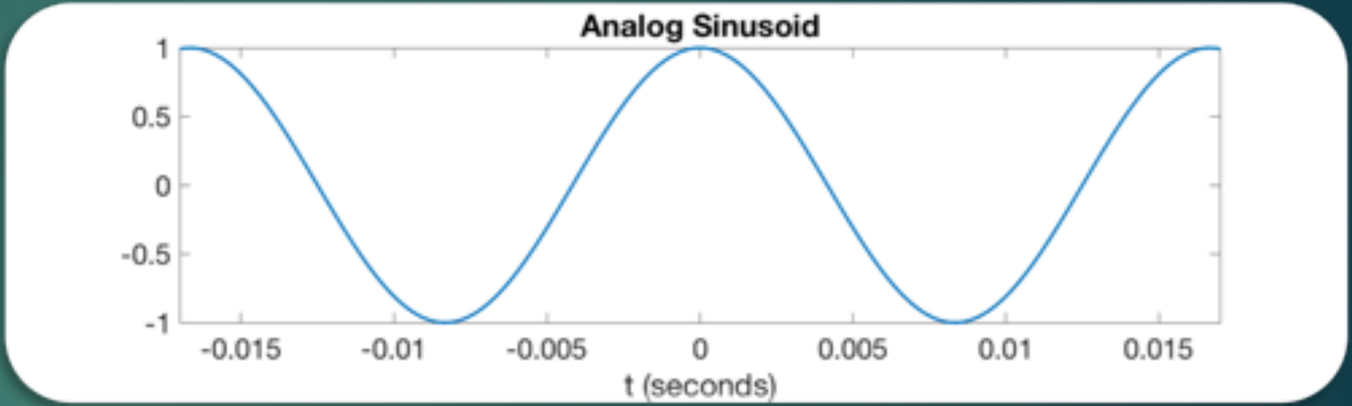
- Let's sample the sinusoid
  - $F_S = 400$ Hz ($T_S = 0.0025$s)



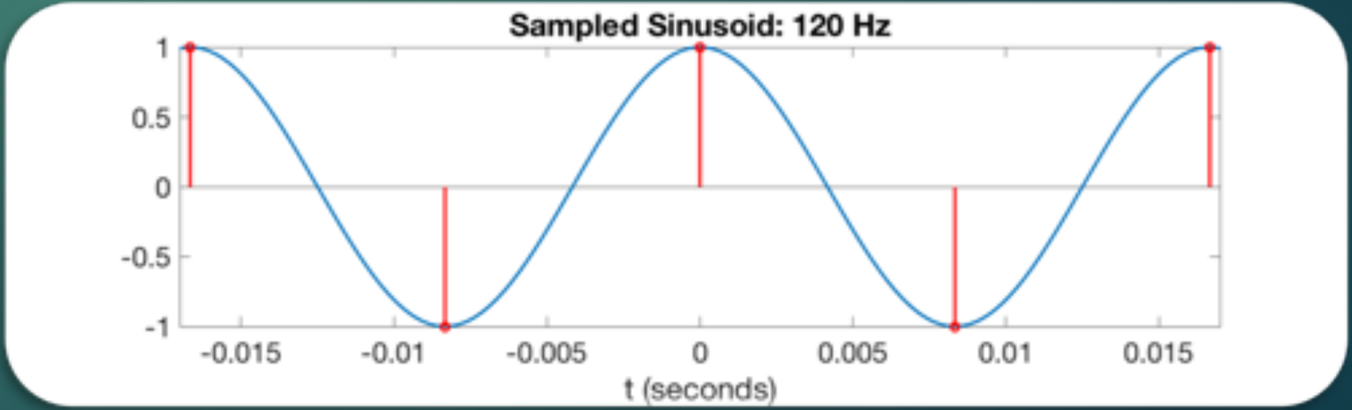
The outline of the sampled signal still matches the analog one

Example: Effects of Aliasing

- Analog Sinusoid
  - x(t) = cos(2πf*t)
  - f = 60 Hz



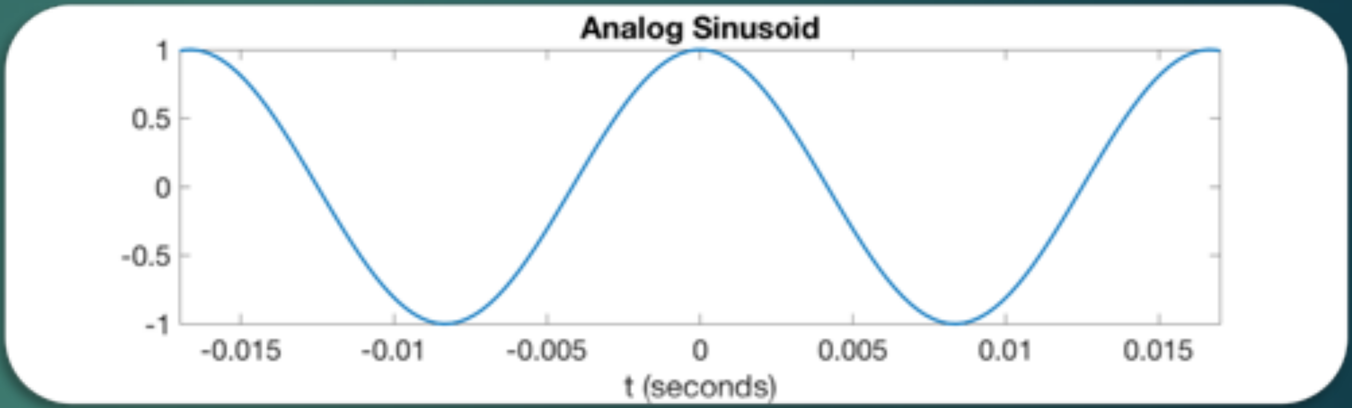
- Let's sample the sinusoid
  - F_S = 120 Hz (Ts = 0.0083s)



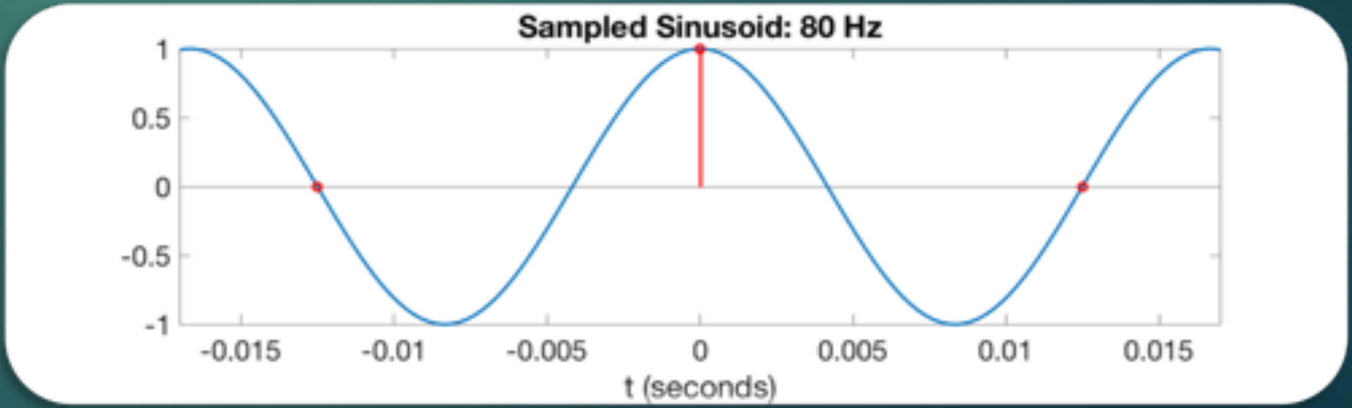
The sampled signal does not reasonable match the analog one

Example: Effects of Aliasing

- Analog Sinusoid
  - x(t) = cos(2πf*t)
  - f = 60 Hz



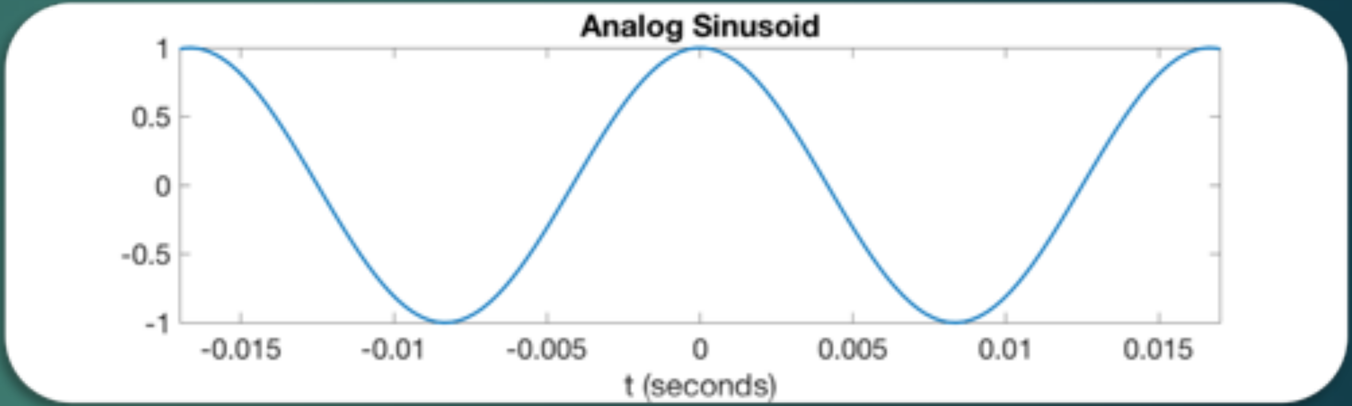
- Let's sample the sinusoid
  - F_S = 80 Hz (Ts = 0.0125)



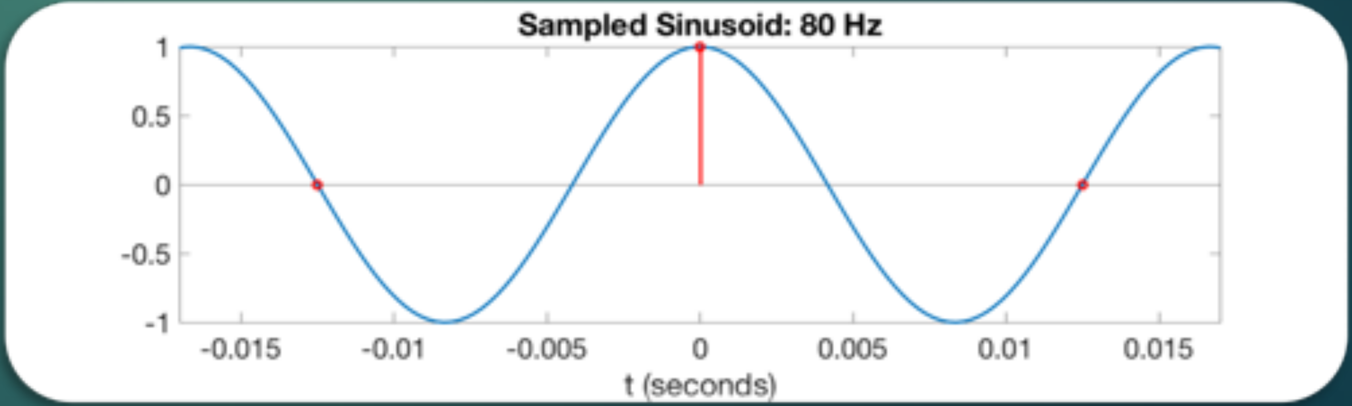
The sampled signal does not match the analog one!

Example: Effects of Aliasing

- Aliasing makes the frequency of the signal appear to be different



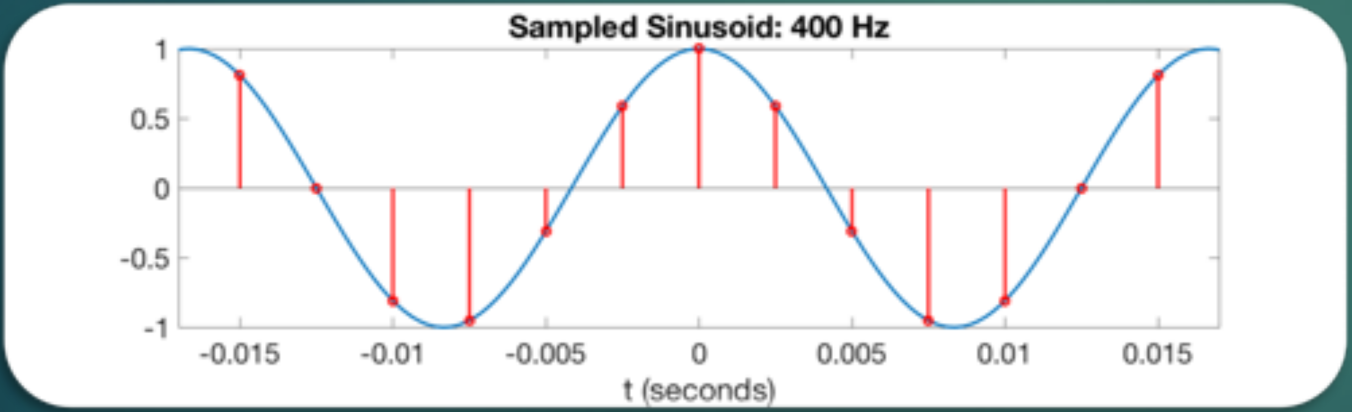
- Aliasing is a ***lossy*** process, meaning samples cannot be recovered once its done.



The sampled signal does not match the analog one!

Avoid Aliasing: The Nyquist Criterion

- To fully represent a signal with a maximum frequency of *f*, use a sampling rate (or frequency) ***> 2f***. This is known as the ***Nyquist Criterion (or Rate)***.



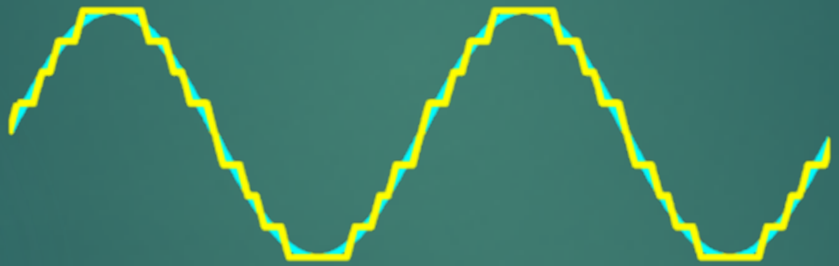
- This signal (blue) is a 60 Hz sine wave, so any sampling rate above 120 Hz is fine.

Learning Objectives

- **You the student will be able to:**
  - Explain how sound is produced from air particles
  - Understand how microphones generate electrical signals from sound particles
  - Describe audio sampling
  - Describe audio quantization

Audio Quantization

- Computers operate on fixed-sized numbers (e.g. 8 – 64 bit)
- Hence, we must quantize the amplitude to a discrete set of values



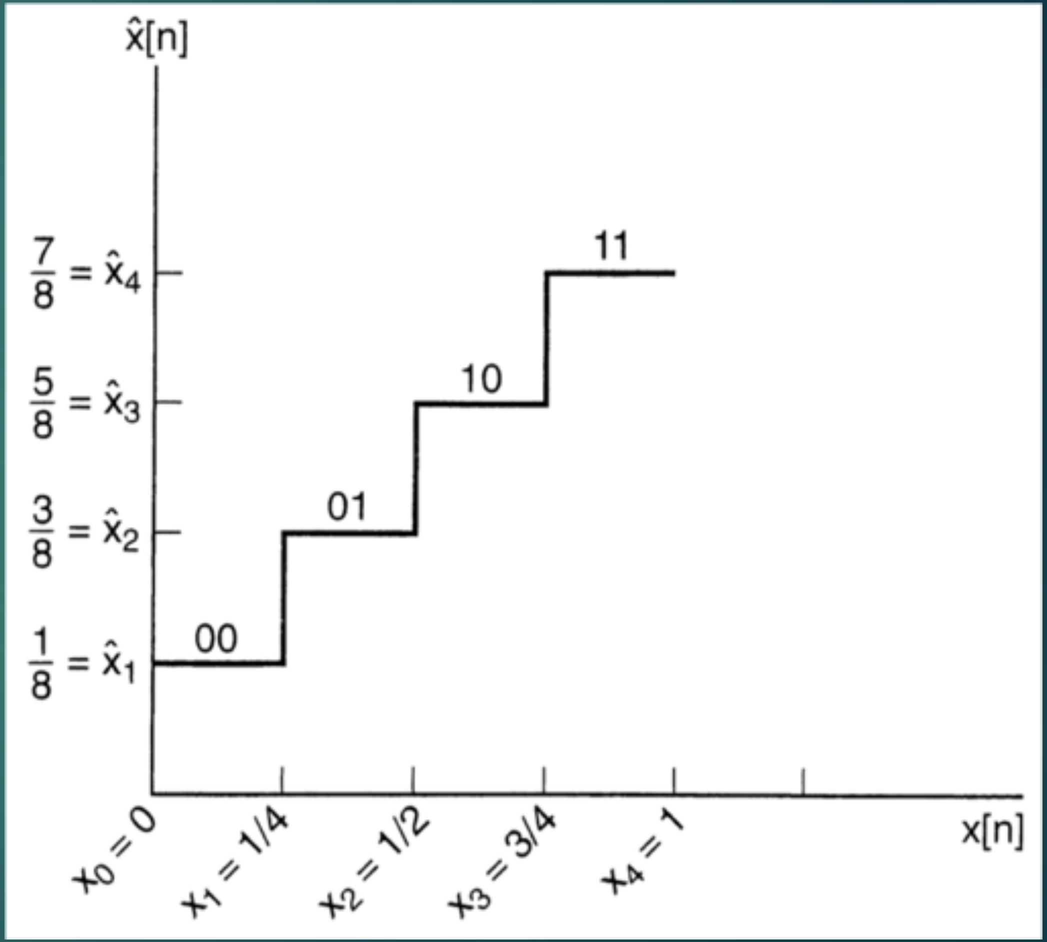
- Let, $\hat{x}[n]$ denote the quantized and discrete (e.g. digital) value
  - If q_i, represents the i^{th} quantization level $(1 \leq i \leq M)$
  - M is the number of quantization levels: (e.g. $M + 1$ decision levels)

$$\hat{x}[n] = q_i, \text{ if } d_{i-1} < x[n] \leq d_i$$

Image courtesy of Prof. Youngmoo Kim, Drexel University

Audio Quantization: Example

- Suppose M = 4
  - *x[n] ranges from [0,1]*
  - *Assume decision and quantization levels are* ***uniform*** *(e.g. equally spaced)*
- *Then,*
  - *Decision levels are* $[0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1]$
  - *Quantization levels* ***can be*** $[\frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}]$
- *An M=4 quantizer can be represented with 2-bit codewords*
  - *Generally, a B-bit binary codebook can represent* 2^B *quantization levels*



Audio Digitalization Specs

- Sampling rate: 44.1 kHz
- Quantization: 16 bits per sample
- Bit rate: 1.411 Mbps (stereo)
  - Bit rate = # bits per sample * F_s

- Sampling rate: 48 kHz (DVD), 96+ kHz (Blu-Ray)
- Quantization: 24 bits per sample
- Bit rate: ?

YouTube Music

- Sampling rate: 44.1+ kHz
- Quantization: 16 or 23 bits per sample

Spotify

- Sampling rate: 44.1+ kHz
- Bit rate: 24 – 320 kbps

SOUNDCLOUD

- Sampling rate: 44.1+ kHz
- Bit rate: 64*- 256 kbps

- Sampling rate: 44.1+ kHz
- Bit rate: 64- 192 kbps

Summary

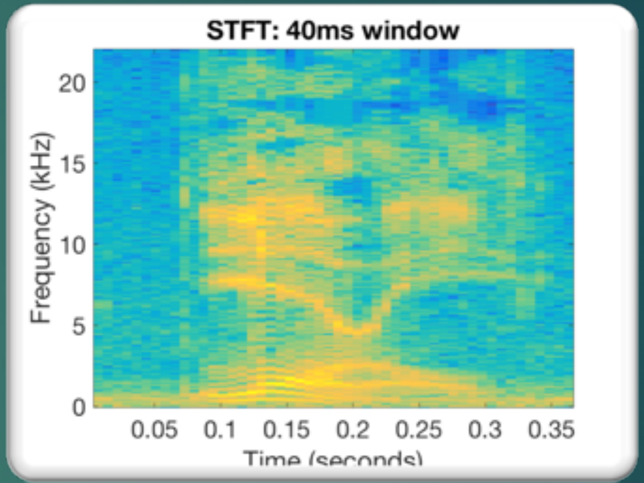
- Sound (and speech) is generated by the displacement (not transition) of air particles
- Sound waves can be captured by microphones, which use diaphragms to capture the amount of particle displacement and convert to an electrical signal
- Computers require that electrical signals are digitized (sampled and quantized)
  - The Nyquist Criterion must be satisfied
  - The number of bits is a system/user preference

In-class assignment

- See handout

Next Class: Speech Processing (Part II)

- Check syllabus on Canvas for pre-class reading assignments
- Topics
  - Time and frequency domain
  - Short-time Fourier Transform
  - Sound level and signal characteristics



$$X[p,k] = \sum_{m=-\infty}^{\infty} x[m]w[pL-m]e^{\frac{-j2\pi km}{N}}$$