



# Experimental Design for Machine Learning on Multimedia Data

## Lecture 7

Prof. Gerald Friedland,  
[fractor@eecs.berkeley.edu](mailto:fractor@eecs.berkeley.edu)

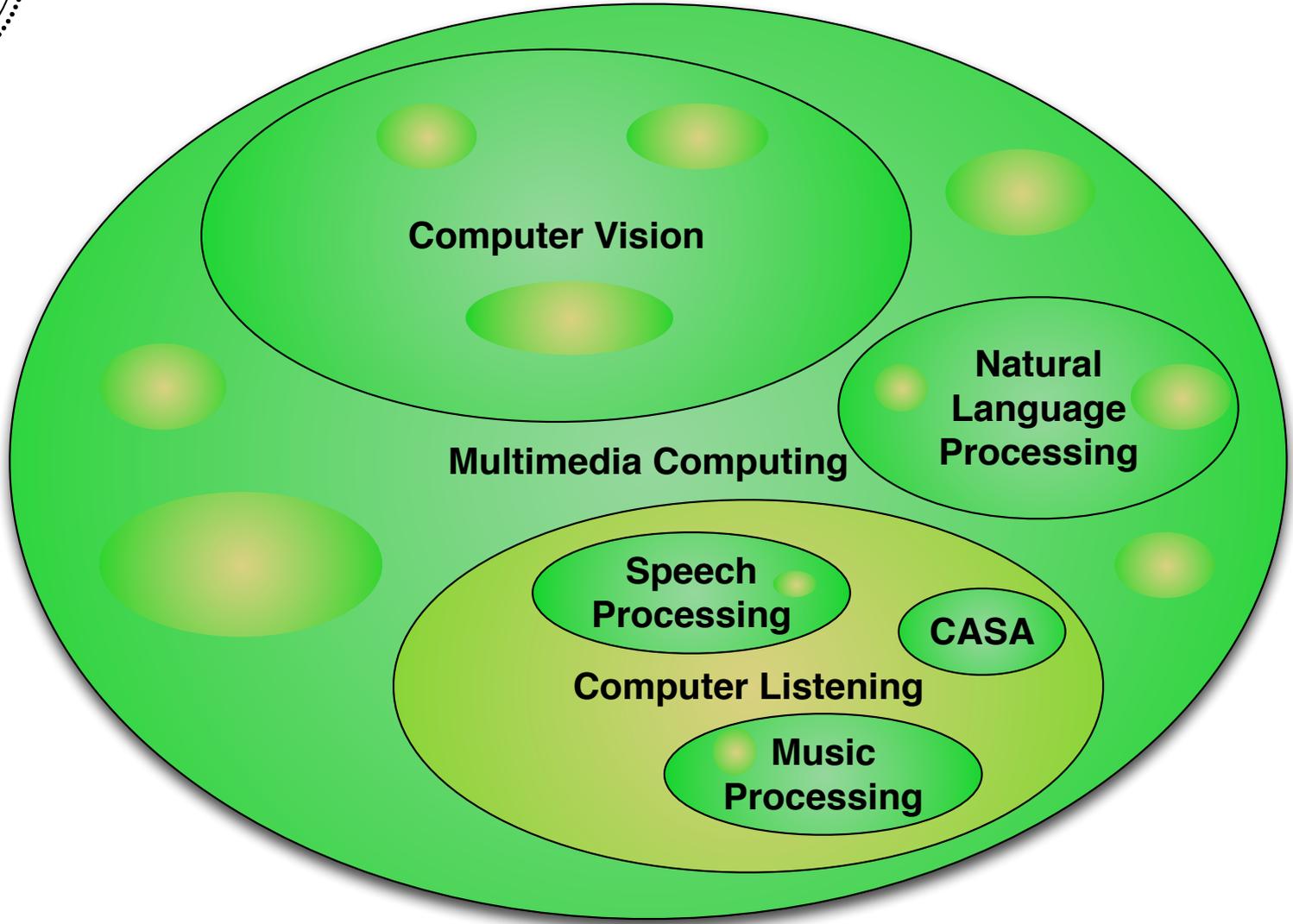


# Today

- Answers to Questions
- Projects
- Sound
  - What is it?
  - How is it recorded and stored?
  - What are its most important properties (to us)?



# Current Situation

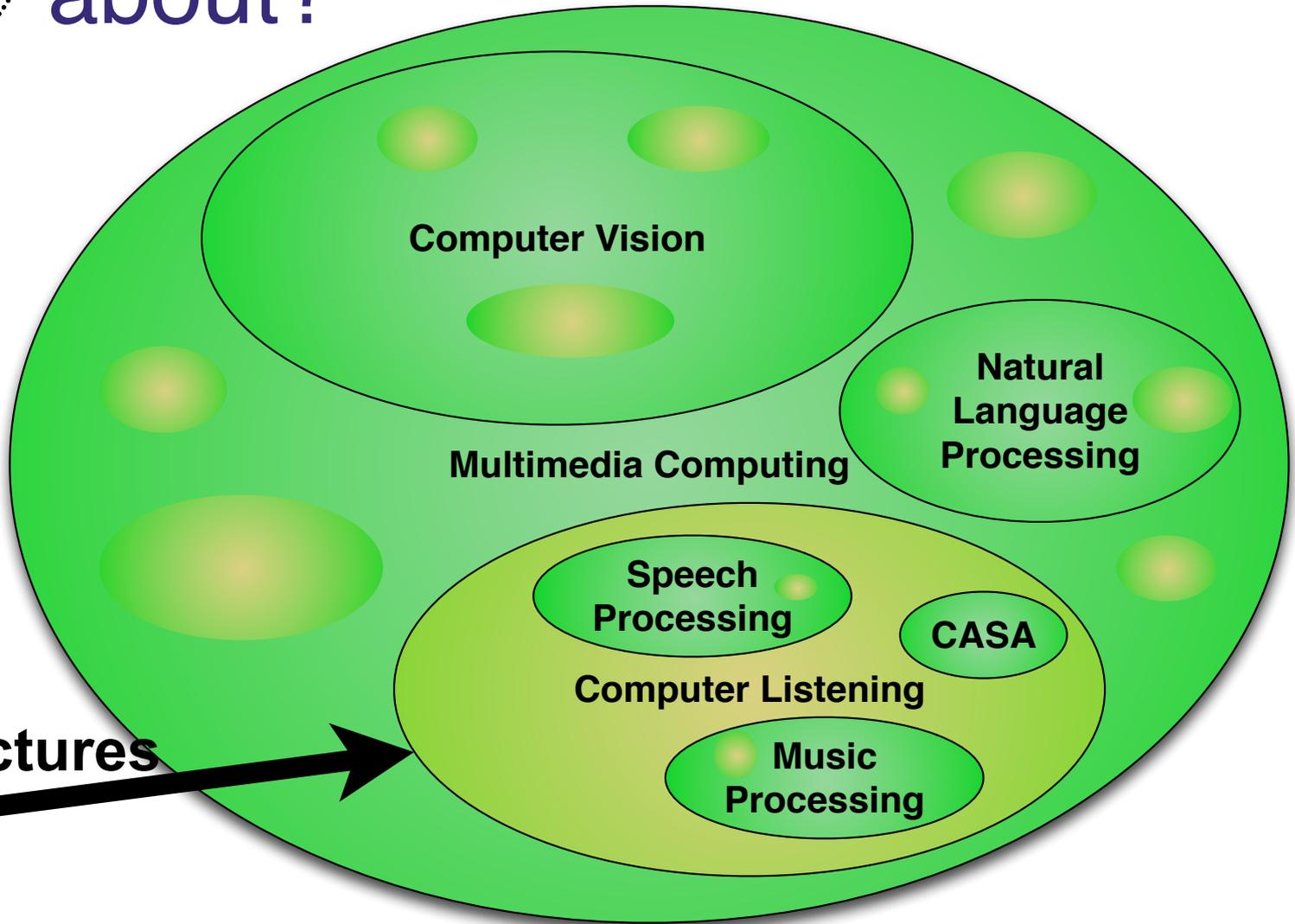


 Area being worked on

 Area not being worked on



# What are the next lectures about?



**Next Lectures**



Area being worked on



Area not being worked on



# Introduction to Sound

- What is sound?
- How is it recorded and stored?
- What are its most important properties (to us)?
- Introduction to features 1
- Some frameworks and tools to work with sound



# What is Sound?

Video from ViHart (Youtube):

[http://www.youtube.com/watch?v=i\\_0DXxNeaQ0](http://www.youtube.com/watch?v=i_0DXxNeaQ0)

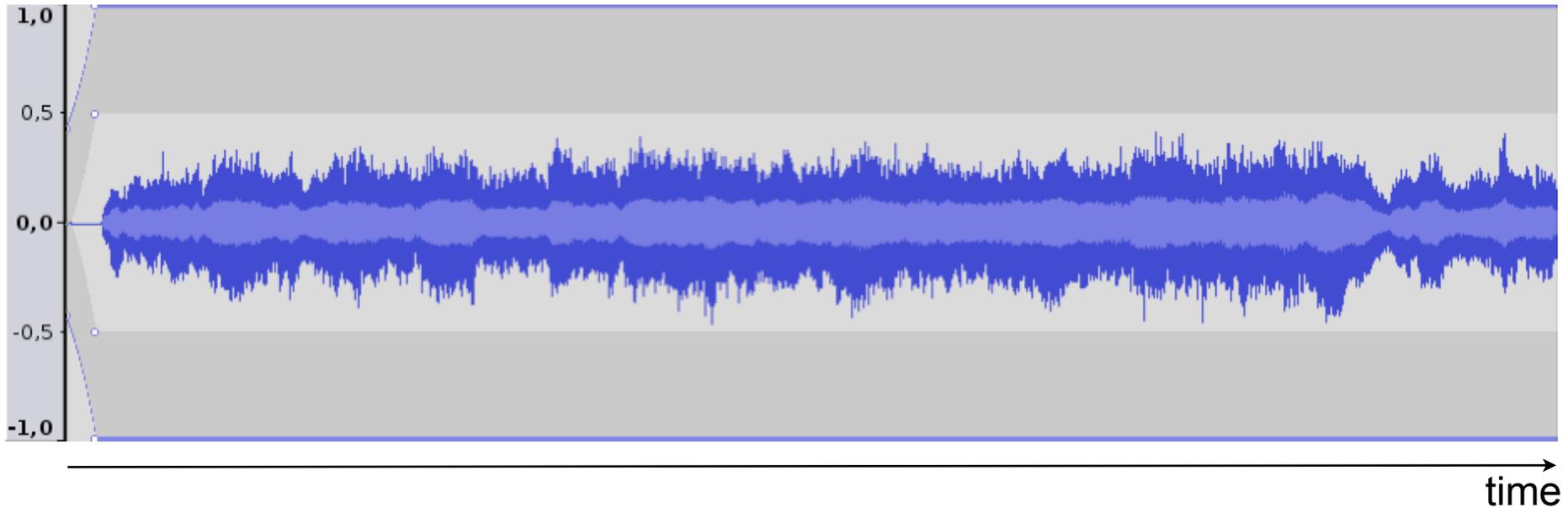


# What is Sound?

“a traveling wave which is an oscillation of pressure transmitted through a solid, liquid, or gas, composed of frequencies within the range of hearing and of a level sufficiently strong to be heard, or the sensation stimulated in organs of hearing by such vibrations.” (AHD)



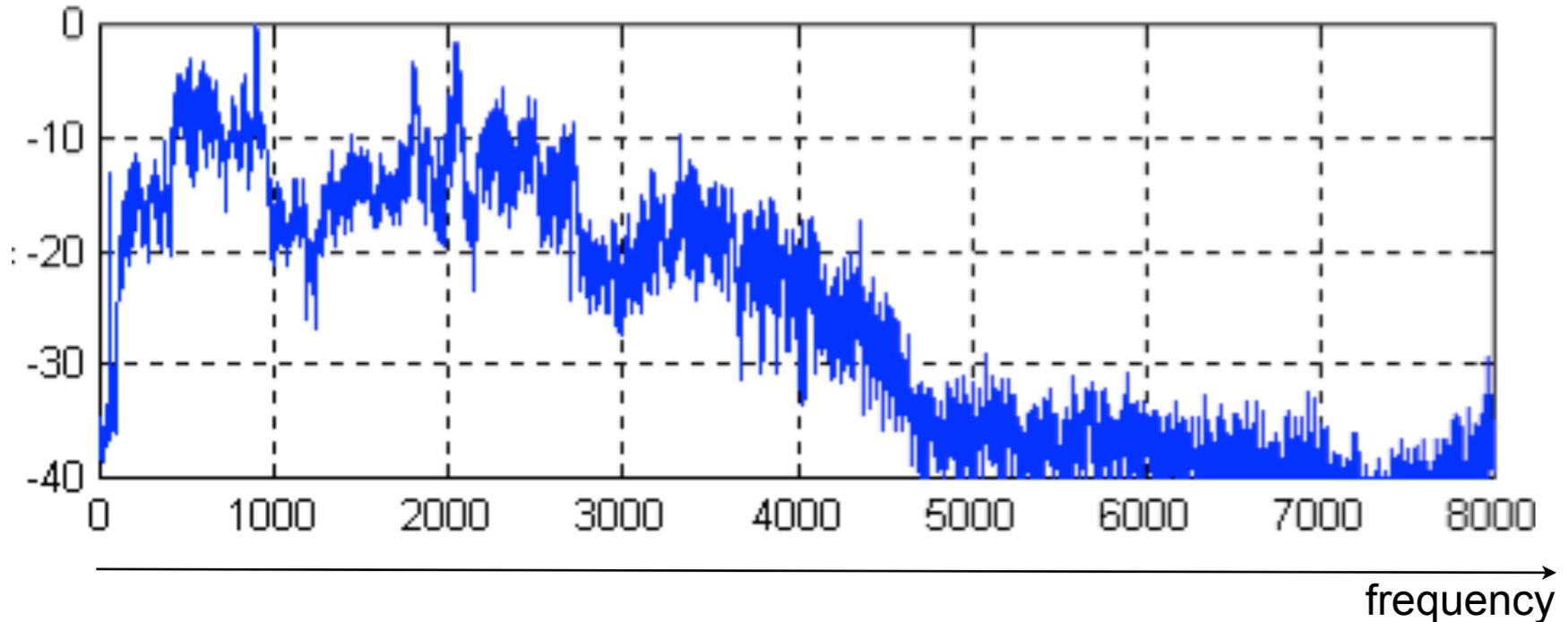
# Visualizations of Sound



**Time Domain aka Amplitude Space aka Waveform**



# Visualizations of Sound



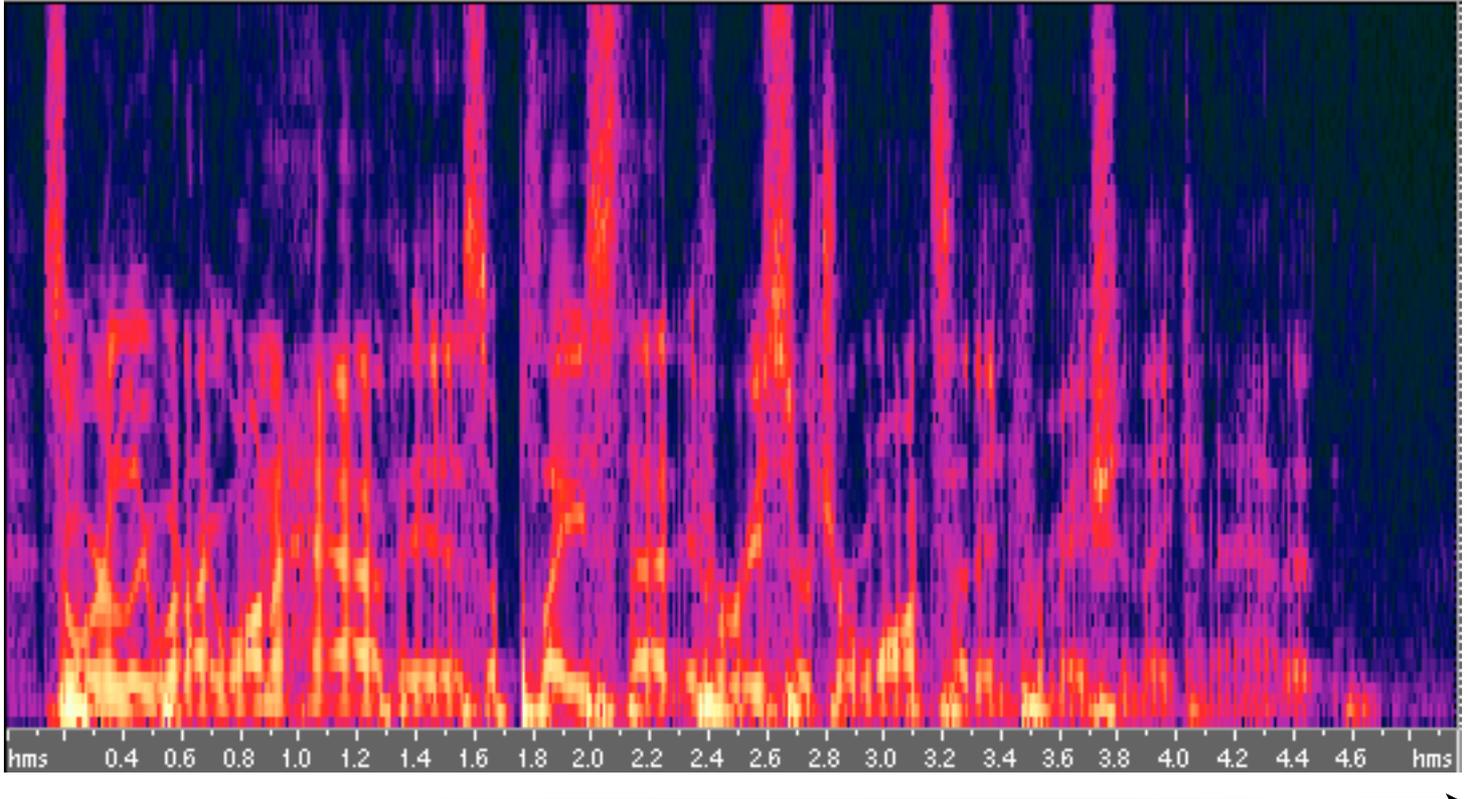
**Frequency Domain aka Fourier Space aka Spectrum**



# Visualizations of Sound

frequency ↑

energy ↙

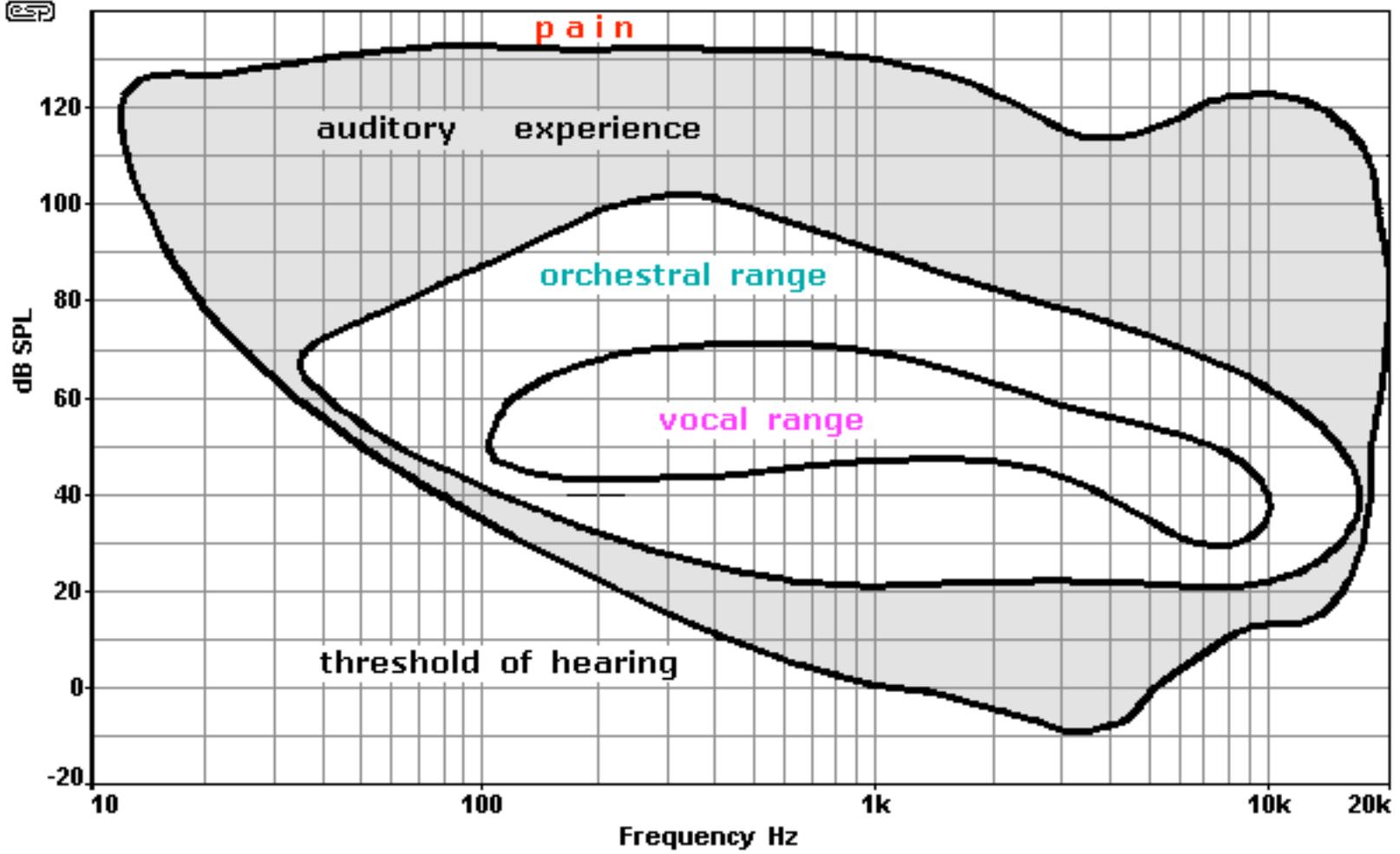


time →

## Spectrogram



# Hearing Spectrum



Source: <http://sound.westhost.com/articles/fadb.htm>



# dB SPL?

- decibel Sound Pressure Level
- NOT a physical unit, only a scale

$$L_p = 10 \log_{10} \left( \frac{p_{\text{rms}}^2}{p_{\text{ref}}^2} \right) = 20 \log_{10} \left( \frac{p_{\text{rms}}}{p_{\text{ref}}} \right) \text{ dB},$$

where  $p_{\text{ref}}$  is the reference sound pressure and  $p_{\text{rms}}$  is the rms sound pressure being measured.

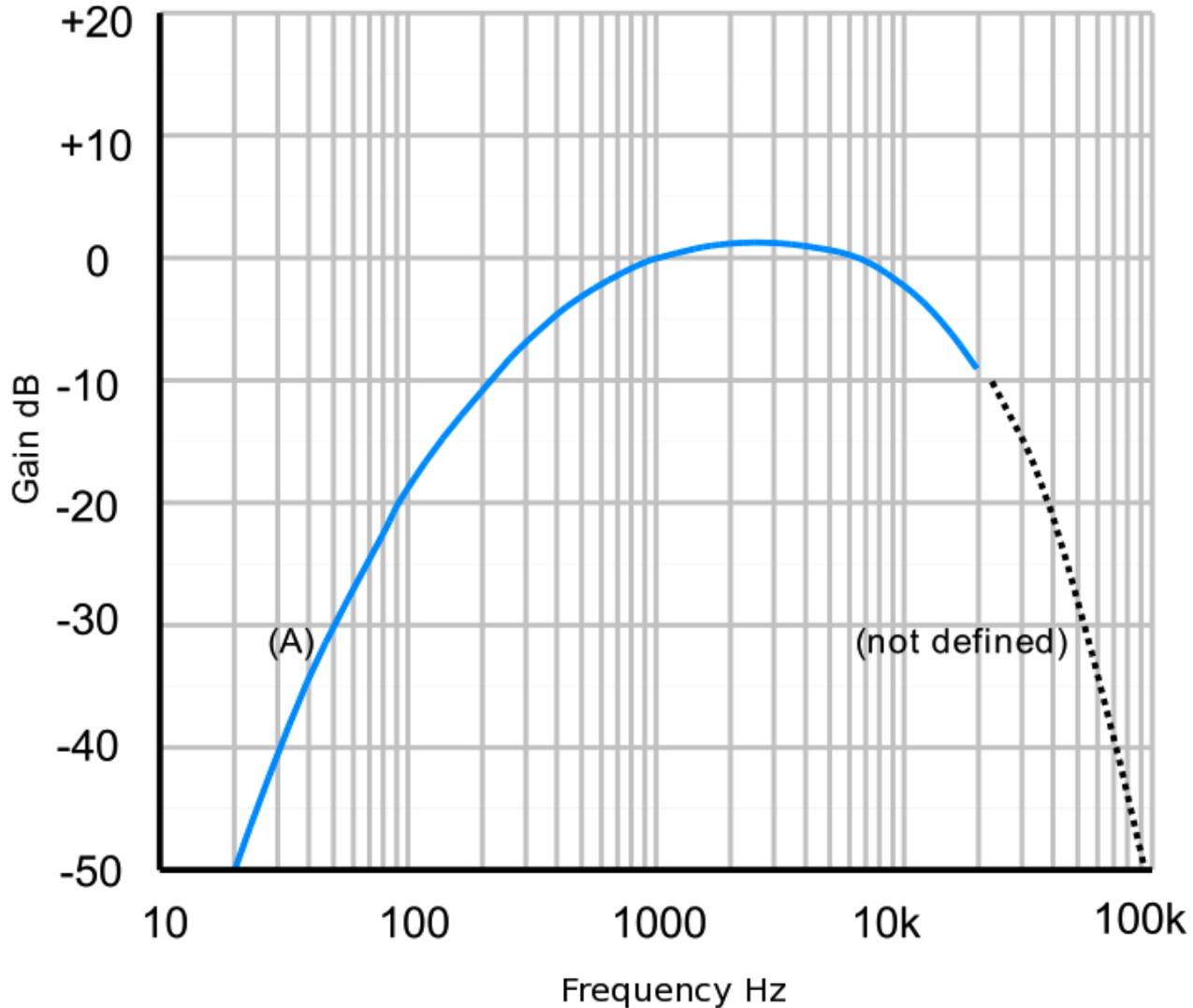


# bits -> db Range (Cheat sheet)

- 8 bits -> 48 dB SPL
- 11 bits -> 66 dB SPL
- 16 bits -> 96 dB SPL
- 24 bits -> 144 dB SPL

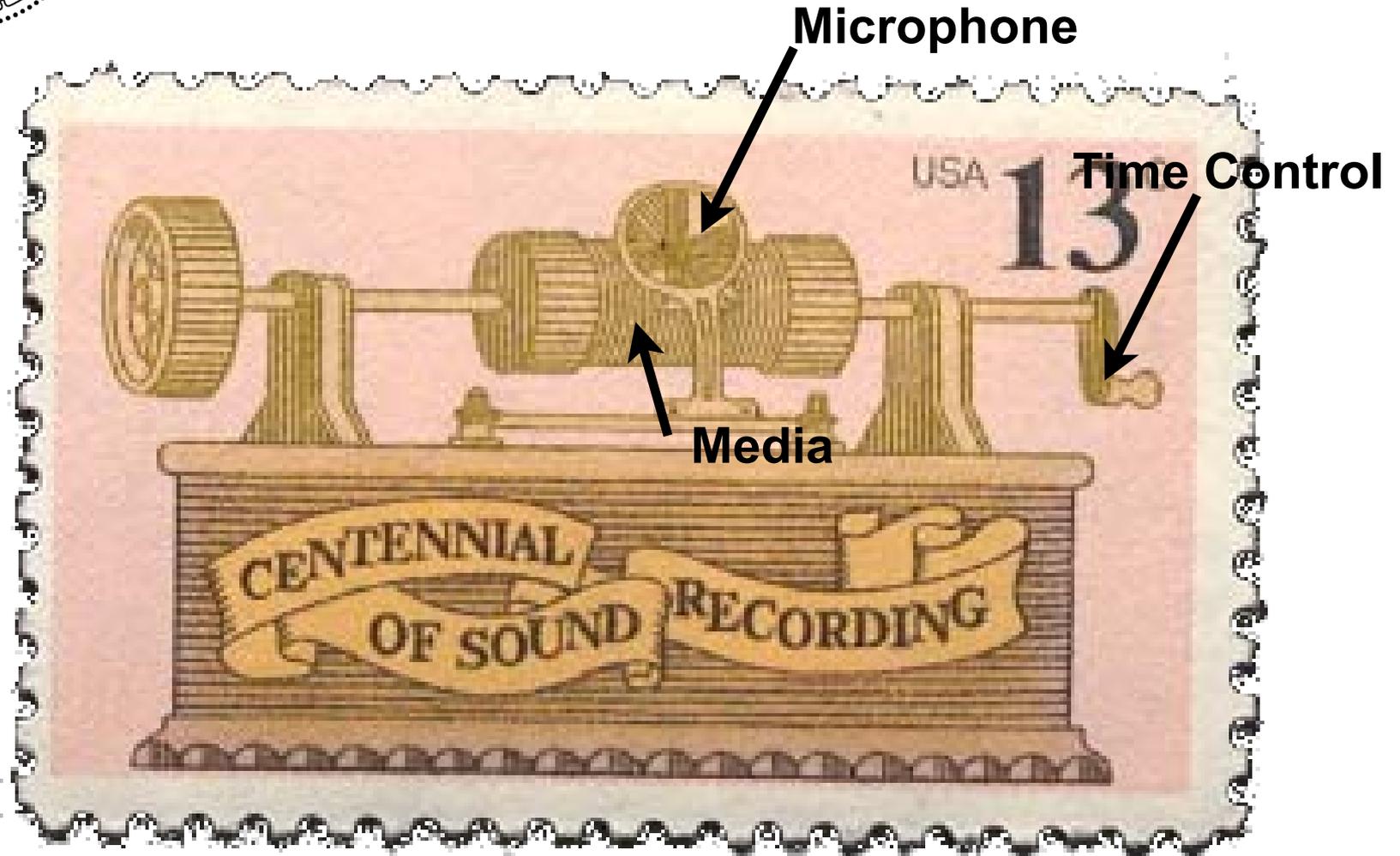


# Frequency-Normalized Range (A-Weighting)





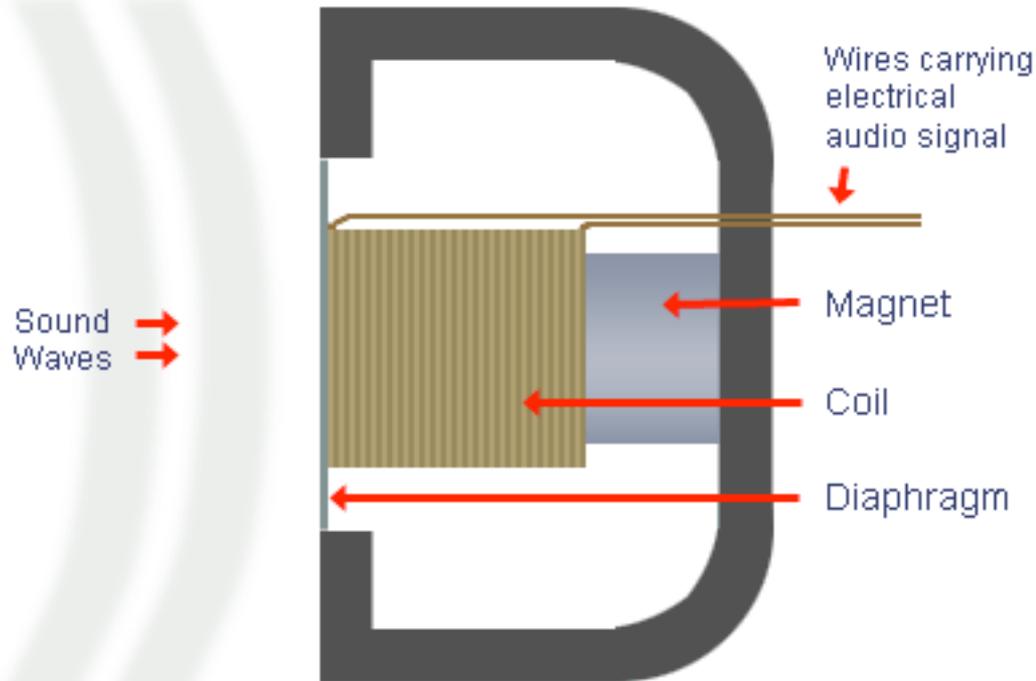
# How is Sound Recorded?





# Modern Microphone

## Cross-Section of Dynamic Microphone



Source: <http://www.mediacollege.com/audio/microphones/dynamic.html>



# Types of Microphones

- Nearfield: Close to sound source e.g., headset, boom microphone (movies, TV productions), singer microphones
- Farfield: Further away from sound source e.g., lapel microphone, stationary microphone, webcams, handheld cams.



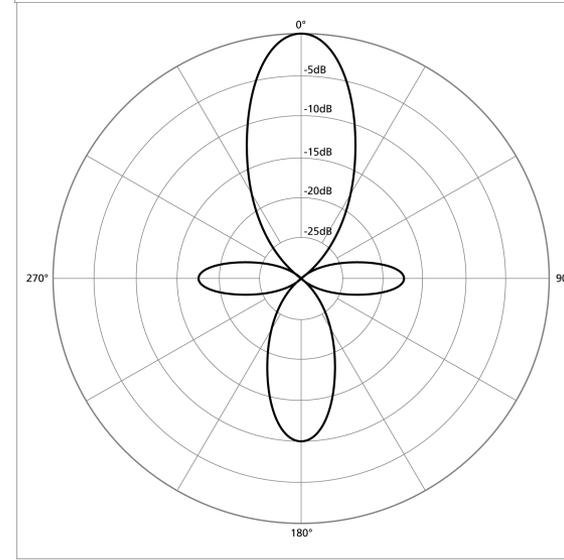
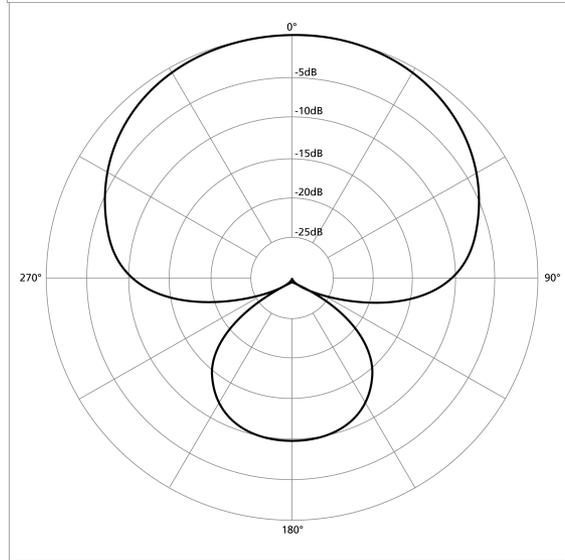
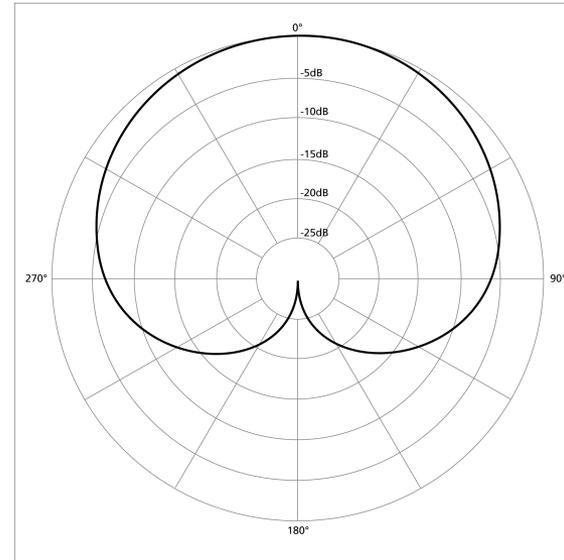
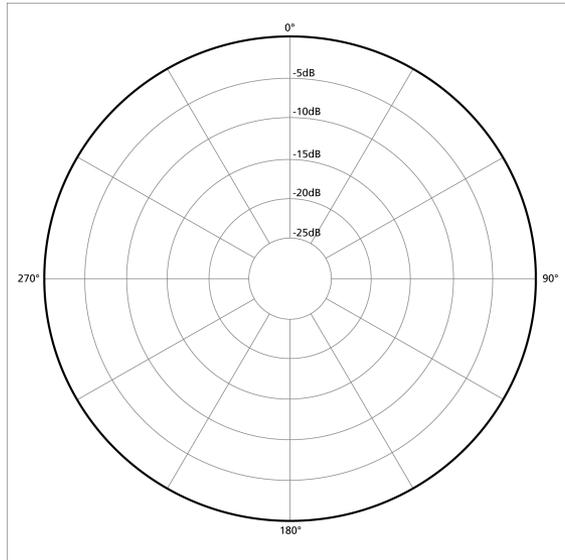
# Difference Farfield/Nearfield

- Nearfield: More energy, less distortion, captures sound source well.
- Farfield: Captures environment with sound source, “better for forensics”, processing often slower.

Demo: <http://www.icsi.berkeley.edu/Speech/mr/nearfar.html>

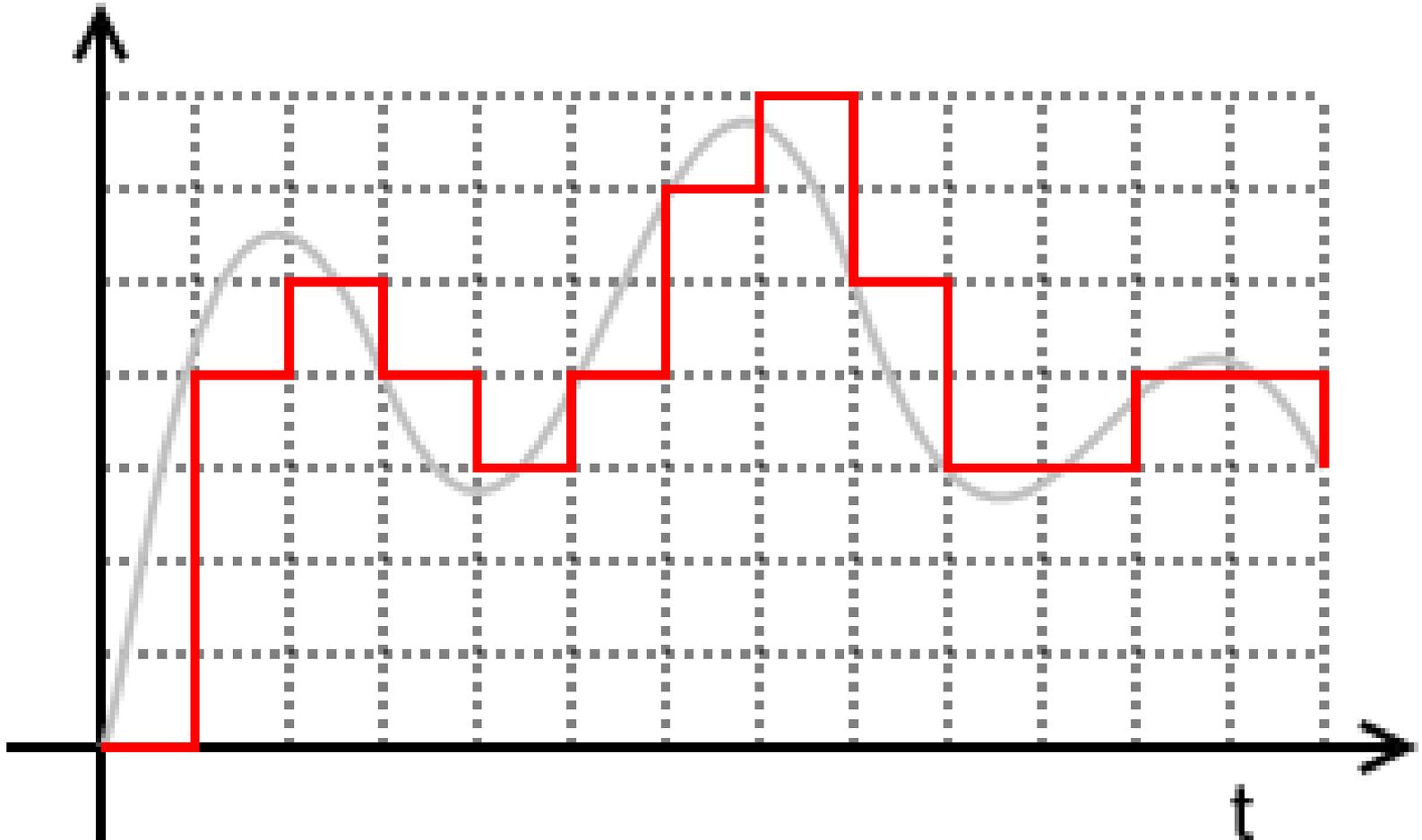


# Microphone Directionality



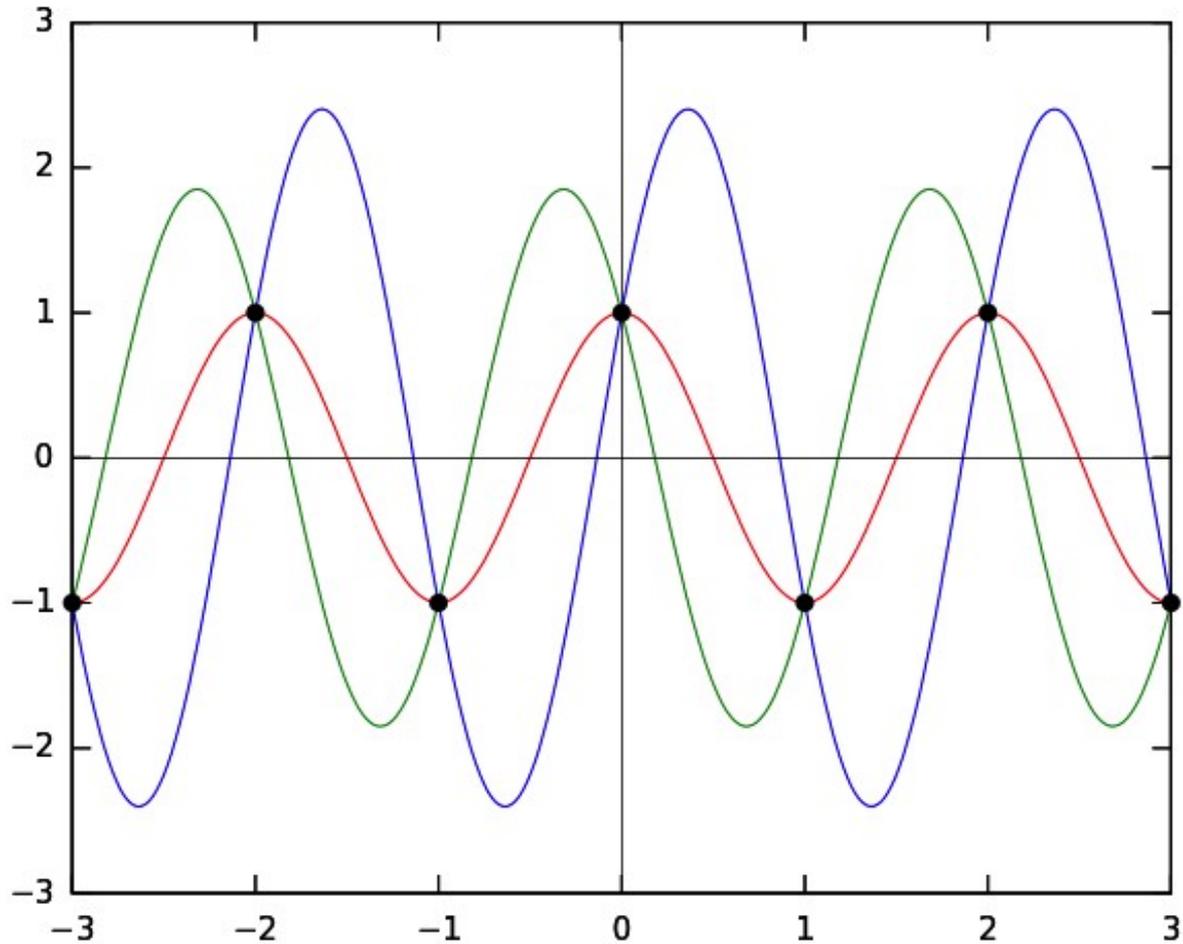


# Digitization of Sound





# Remember: Nyquist Limit!



Math: See Draft Chapter 3 of Friedland & Jain on [mm-creole.org](http://mm-creole.org)

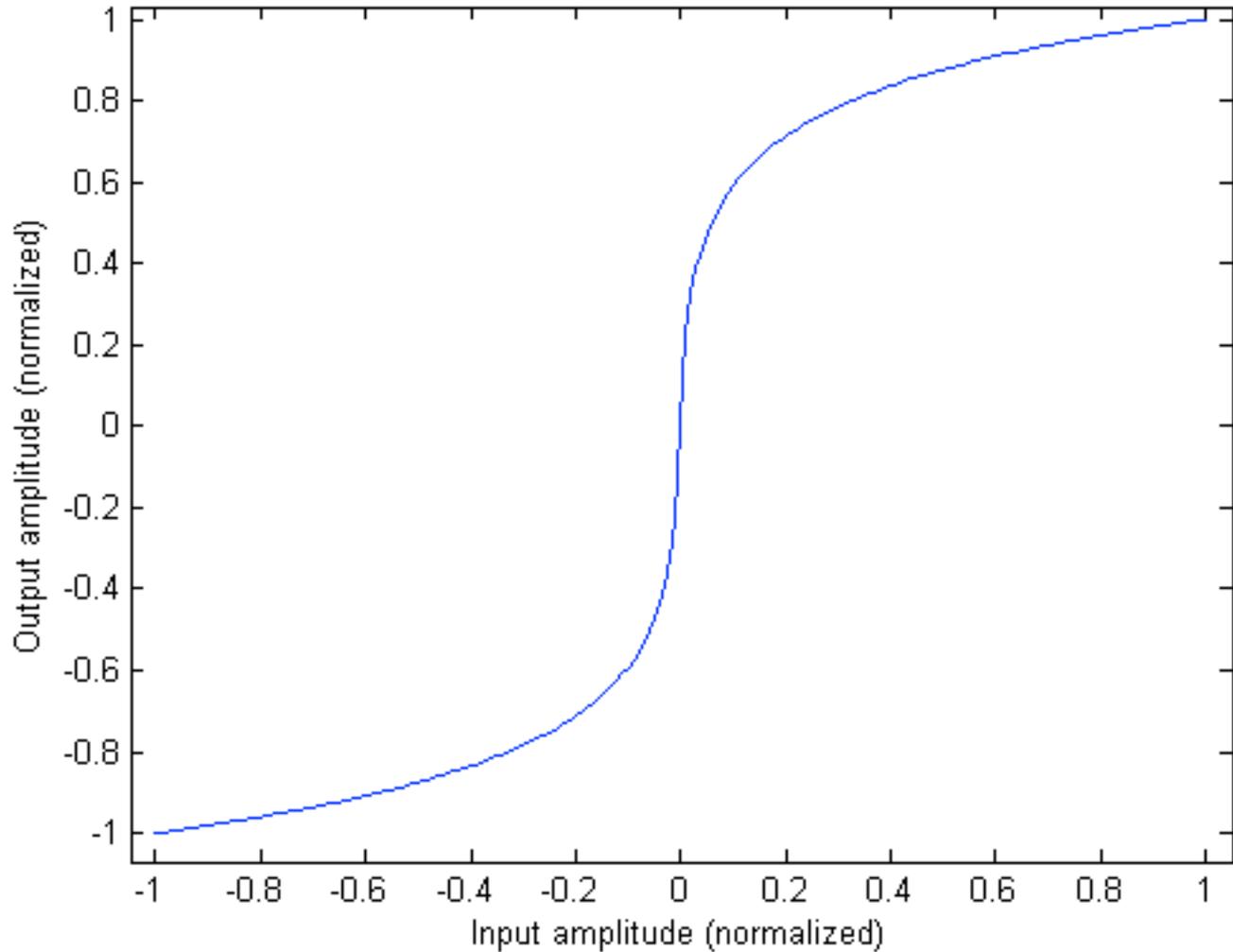


# Common Recording Resolutions

- 8000Hz, 8-bit log. companded ~ 11 bit uncompanded (a/ $\mu$ -law): telephone
- 16000Hz, 16-bit linear: speech (Skype)
- 44100Hz, 16-bit linear, stereo: Compact Disk, many camcorders
- 48000Hz, 32-bit linear, stereo: Digital Audio Tape, Hard Disk Recorders



# $\mu$ -law Compressing





# Next Week

- Intro to Audio Analysis