Experimental Design for Machine Learning on Multimedia Data
Lecture 8

Prof. Gerald Friedland,
fractor@eecs.berkeley.edu

Website: http://www.icsi.berkeley.edu/~fractor/fall2019/

Today

- Time plan: Remaining Semester
- Reproducibility vs Repeatability
- Sound
  - What is it?
  - How is it recorded and stored?
  - What are it’s most important properties (to us)?

Timeplan

- November 8 (today): Audio 1
- November 15th: Audio 2 & Vision
- November 22nd: Project Presentations 1
- November 29th: Project Presentations 2
- December 6th: Project Presentations 3
- (December 13th: Project Presentations 4)
- December 20th: Final
- December 20th: Project deadline (TBD)

Today

- Reproducibility vs Repeatability

- Sound
  - What is it?
  - How is it recorded and stored?
  - What are it’s most important properties (to us)?

Thought Framework: Reproducibility

An experimental result is not fully established unless it can be independently reproduced. ***https://project.inria.fr/acmmmreproducibility/***

- Repetition of experiment: Predict outcome X, observe outcome X.
- Reproduction of an experiment: Independent party recreates the experimental setup of an experiment and repeats it.
- Reproduction requires a complete description of all factors that make the experiment repeatable.
- Understanding: What are the minimum amount of factors that make the experiment repeatable and how do they influence each other?

Repetition, Reproduction, Understanding?

- Downloading the docker instance an re-running on your campus.
- Downloading the trained model and re-running it on the same data.
- Using the exact software versions and exact configuration of hyper parameters and retraining the same model to obtain similar results.
- Taking the description from a paper, rebuilding the setup as described and obtaining the same accuracy and adversarial examples.
- Taking the description from a paper, rebuilding the setup as described, obtaining the same accuracy and adversarial examples and explaining why a different setup yields other limits.

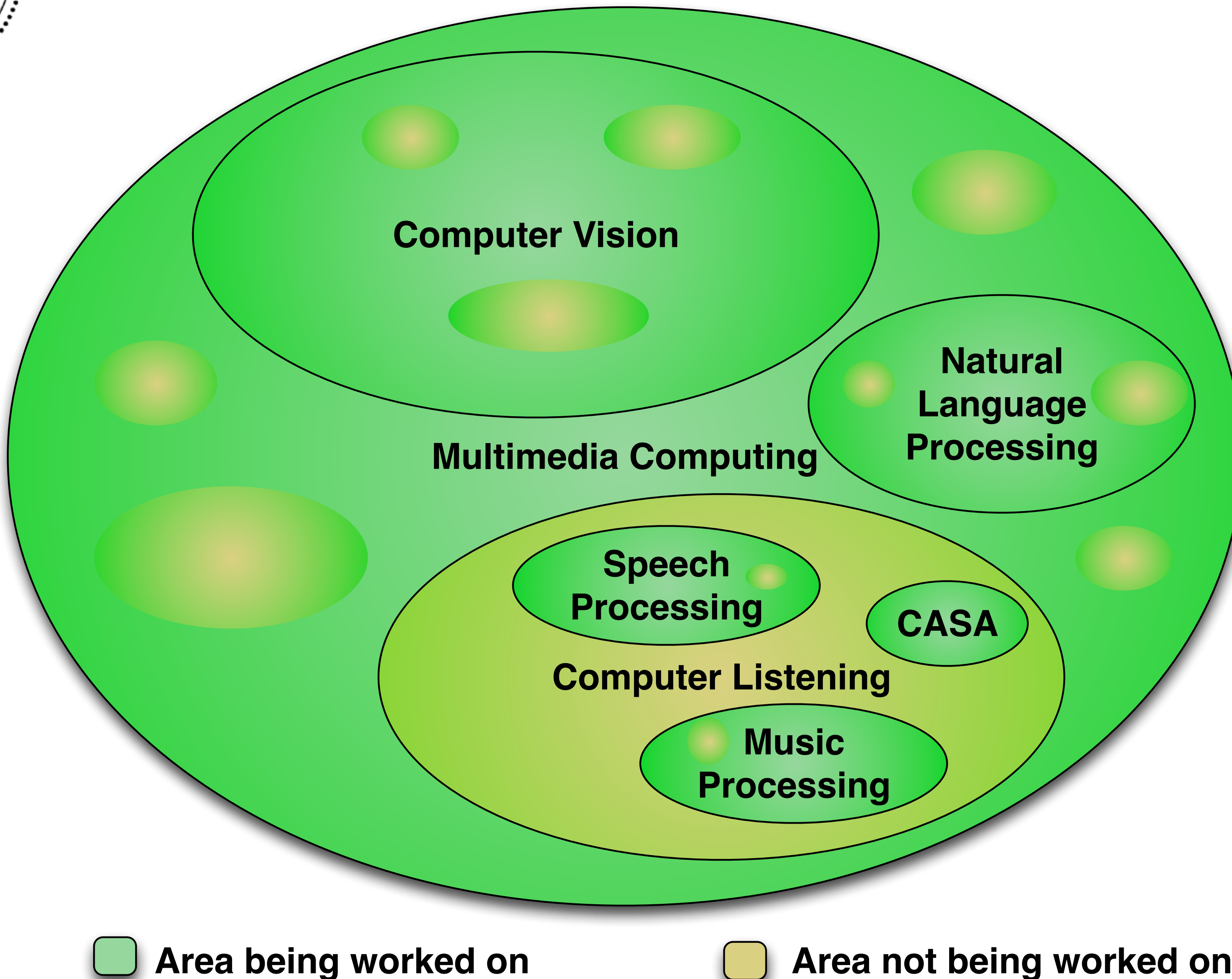
What factors make Machine Learning Repeatable, Reproducible, Understandable?

- Repeatable:
  Set of all hyper parameters, seed, architecture, exact software versions of all libraries, exact order of training sample presentation, etc (?)

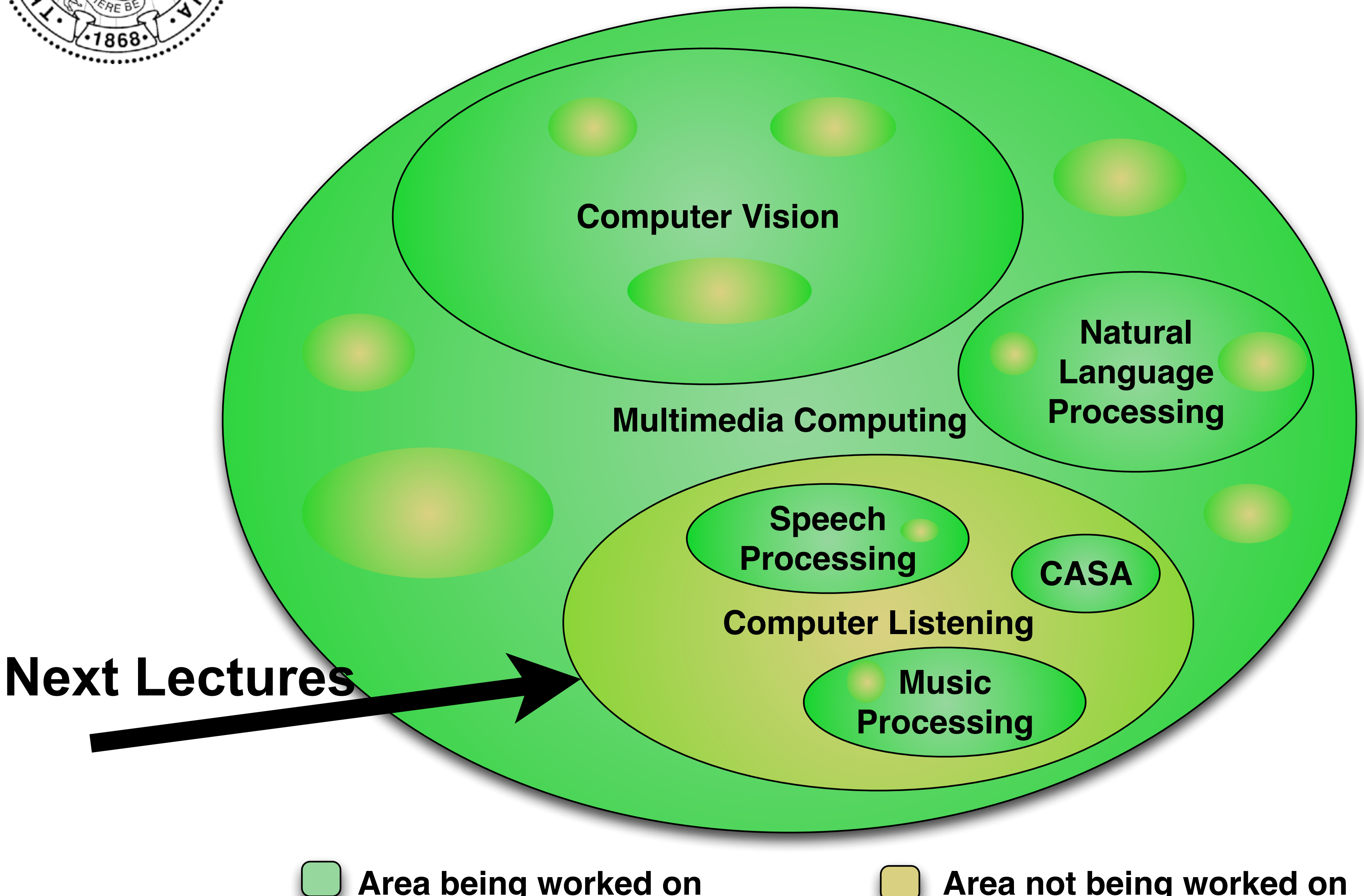
- Reproducible:
  Capacity, perfect training (i.e. training that guarantees to reach a global minimum error).

- Understandable:
  Minimum capacity, perfect training.

Current Situation



What is this lecture about?



Introduction to Sound

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

What is Sound?

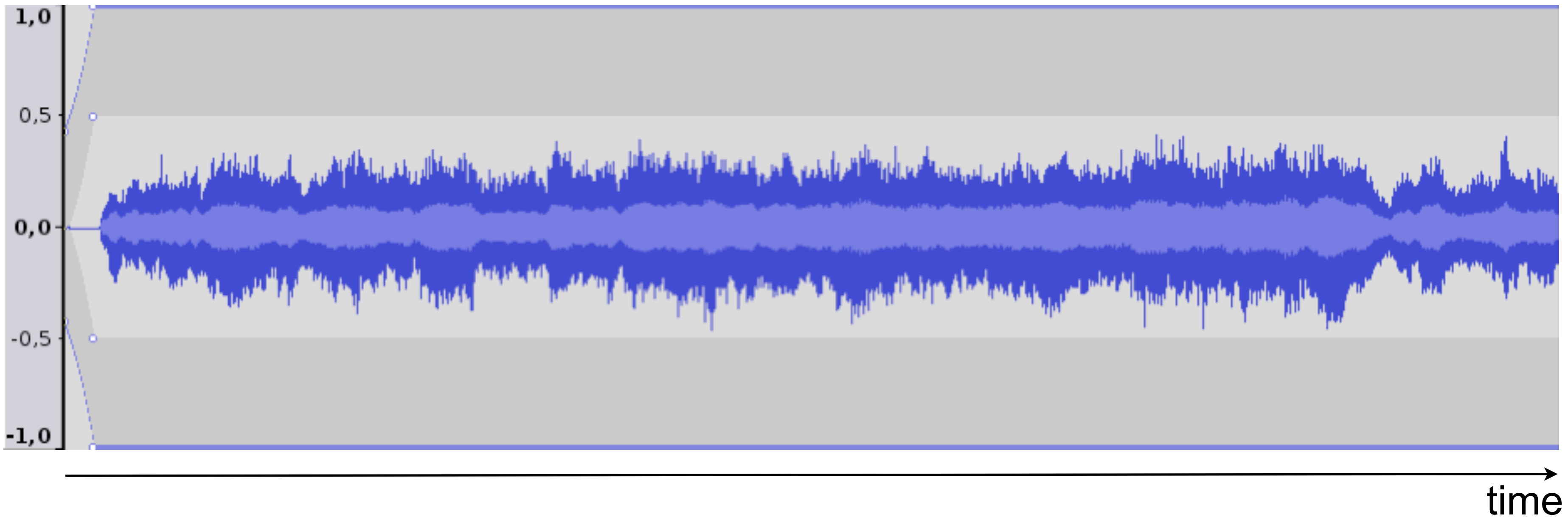
Video from ViHart (Youtube):

http://www.youtube.com/watch?v=i_0DXxNeaQ0

What is Sound?

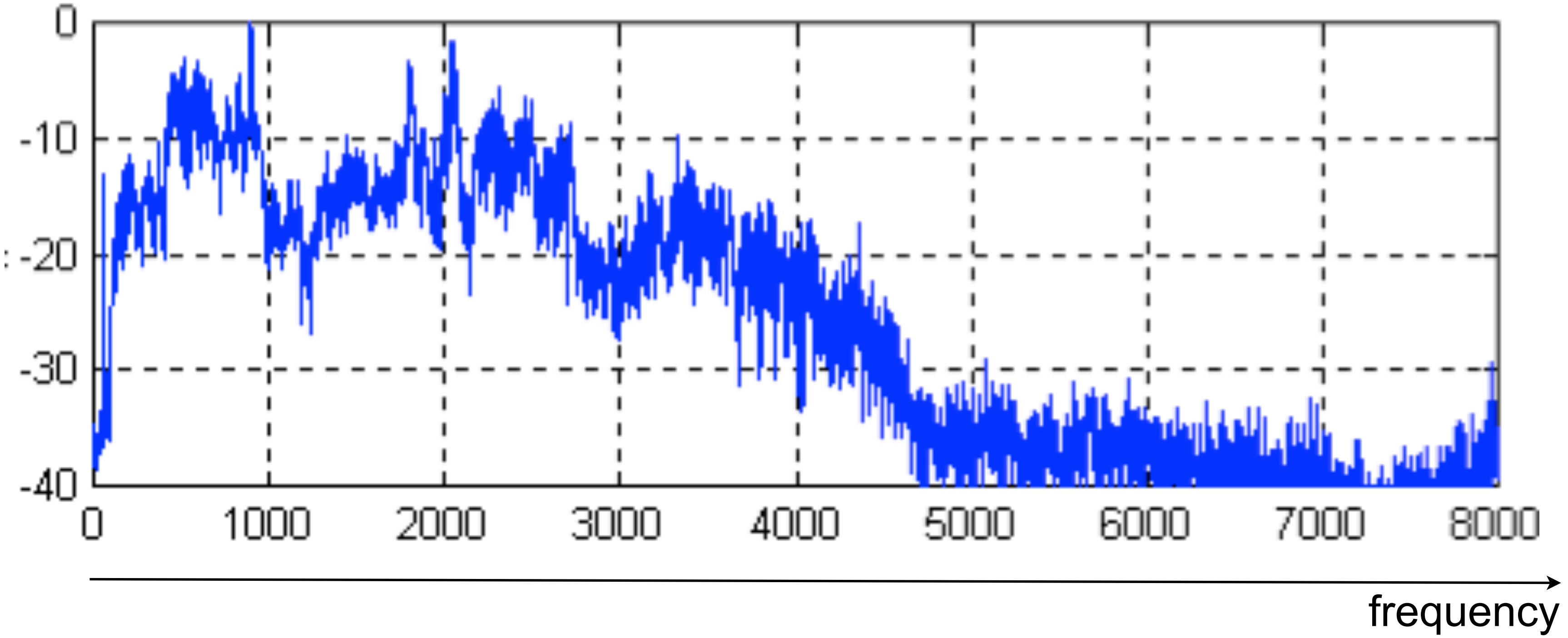
“a traveling wave which is an oscillation of pressure transmitted through a solid, liquid, or gas, composed of frequencies within the range of hearing and of a level sufficiently strong to be heard, or the sensation stimulated in organs of hearing by such vibrations.” (AHD)

Visualizations of Sound



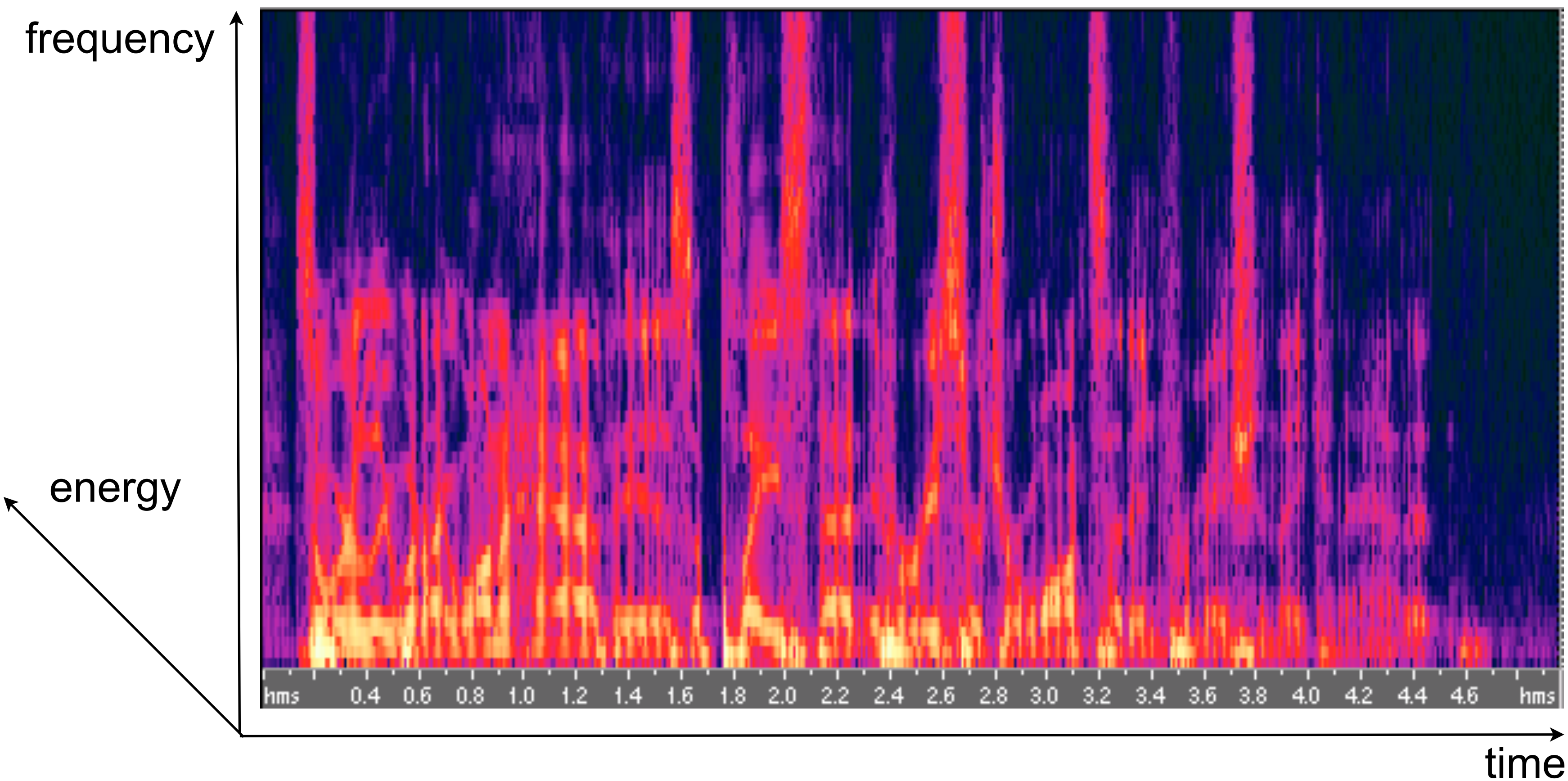
Time Domain aka Amplitude Space aka Waveform

Visualizations of Sound



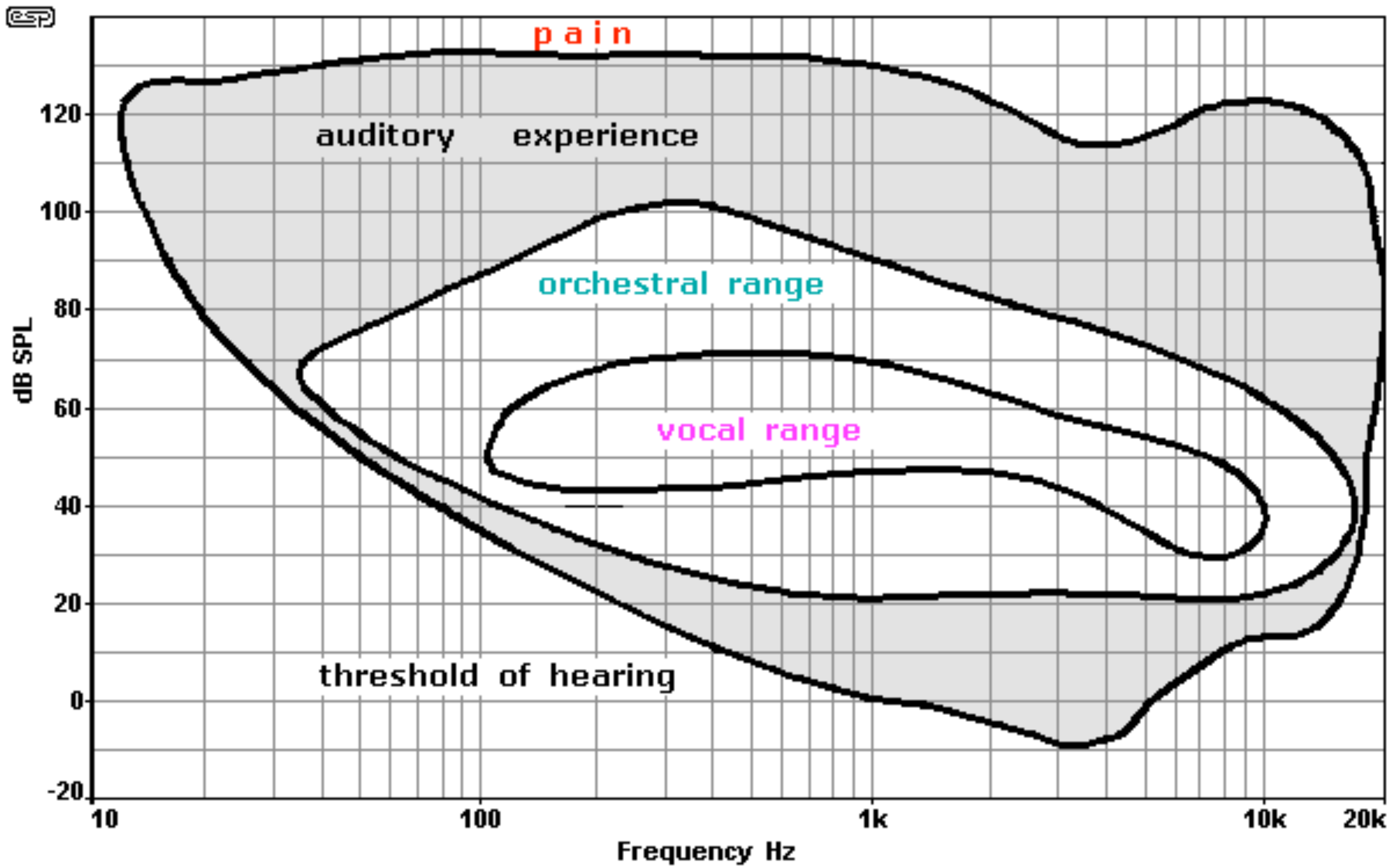
Frequency Domain aka Fourier Space aka Spectrum

Visualizations of Sound



Spectrogram

Hearing Spectrum



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

dB SPL?

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

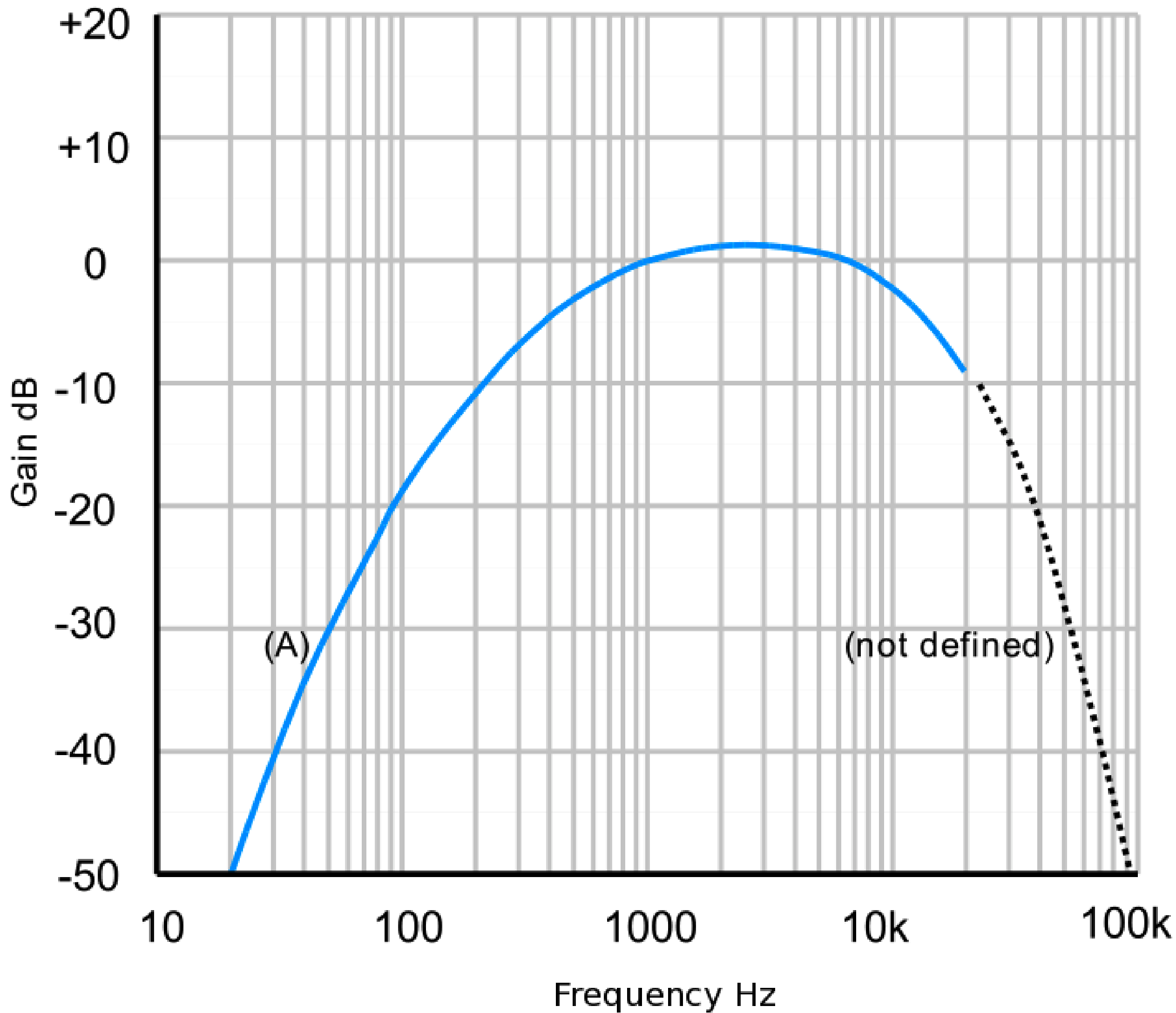
$$L_p = 10\log_{10}\left(\frac{{p_{\mathrm{rms}}}^2}{{p_{\mathrm{ref}}}^2}\right) = 20\log_{10}\left(\frac{p_{\mathrm{rms}}}{p_{\mathrm{ref}}}\right)\ \mathrm{dB},$$

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

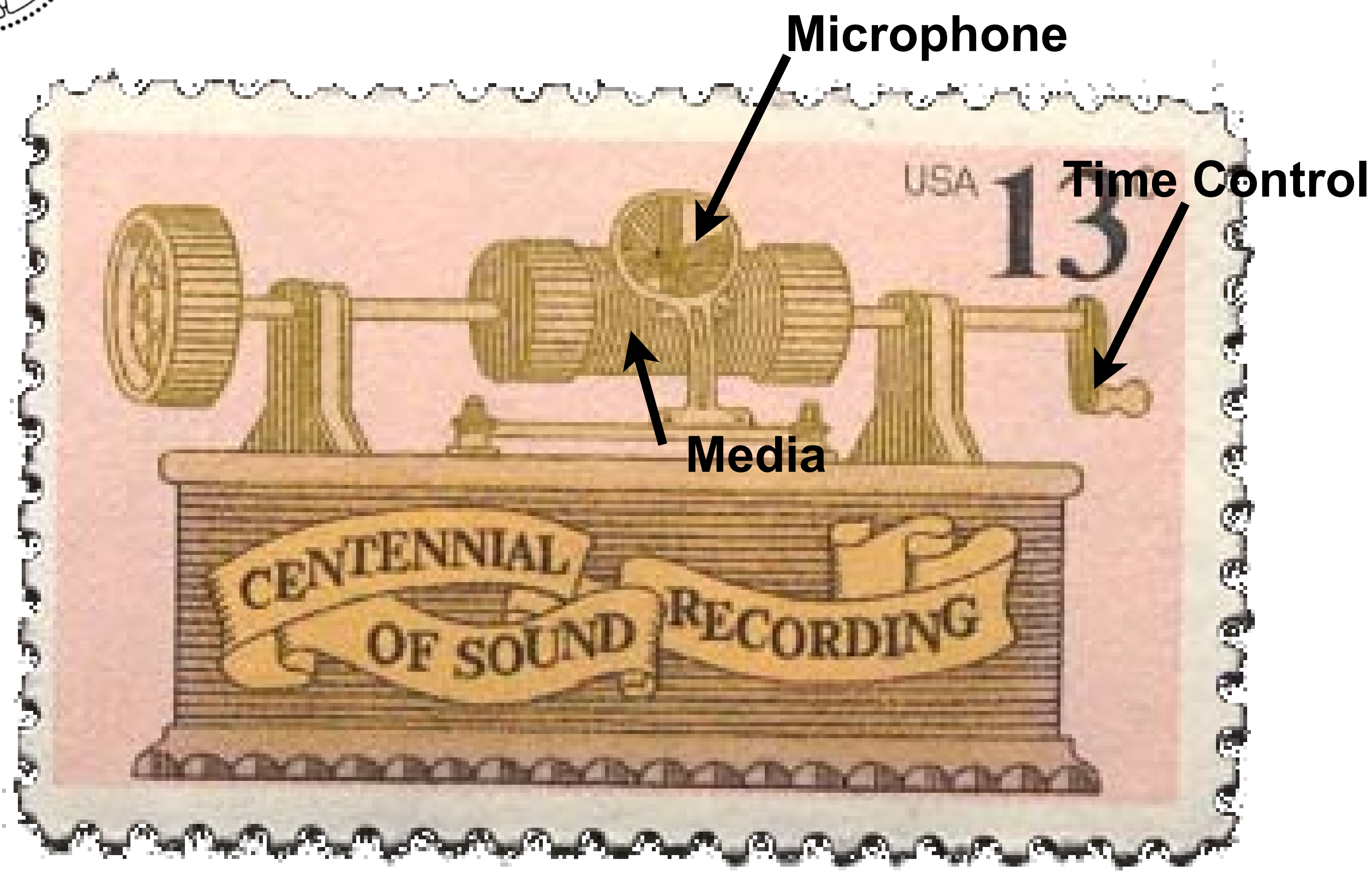
bits -> db Range (Cheat sheet)

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

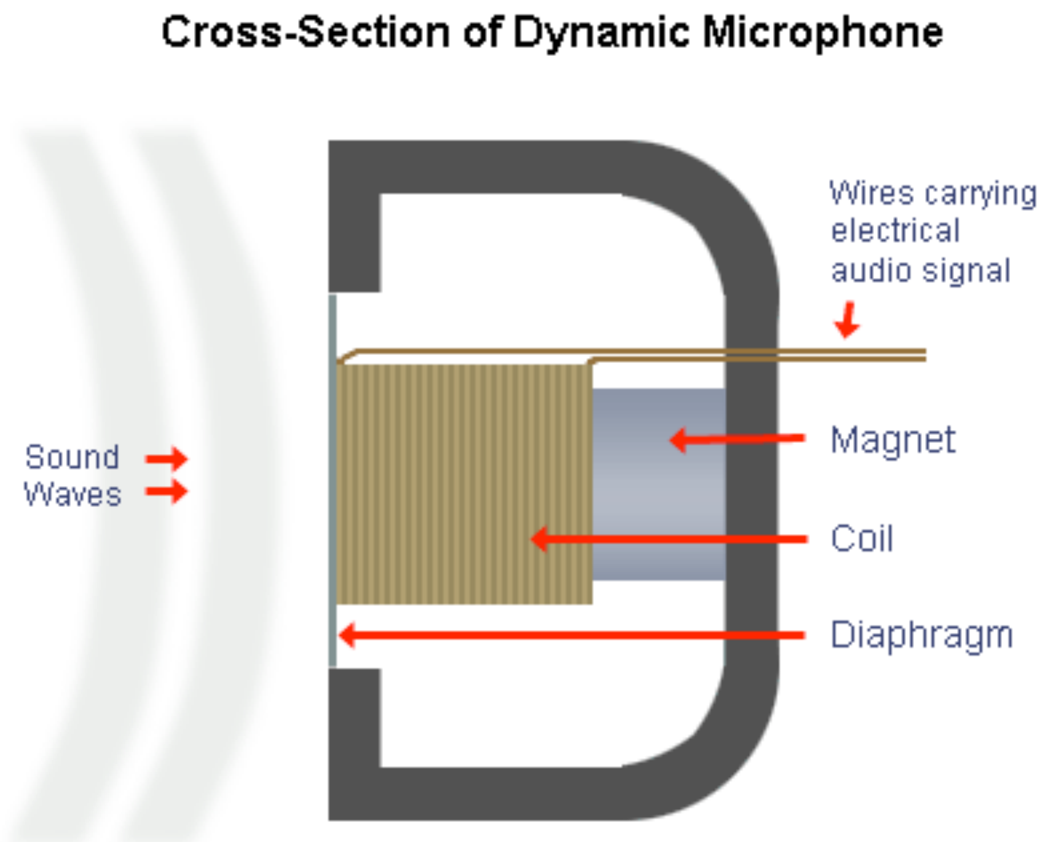
Frequency-Normalized Range (A-Weighting)



How is Sound Recorded?



Modern Microphone



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.

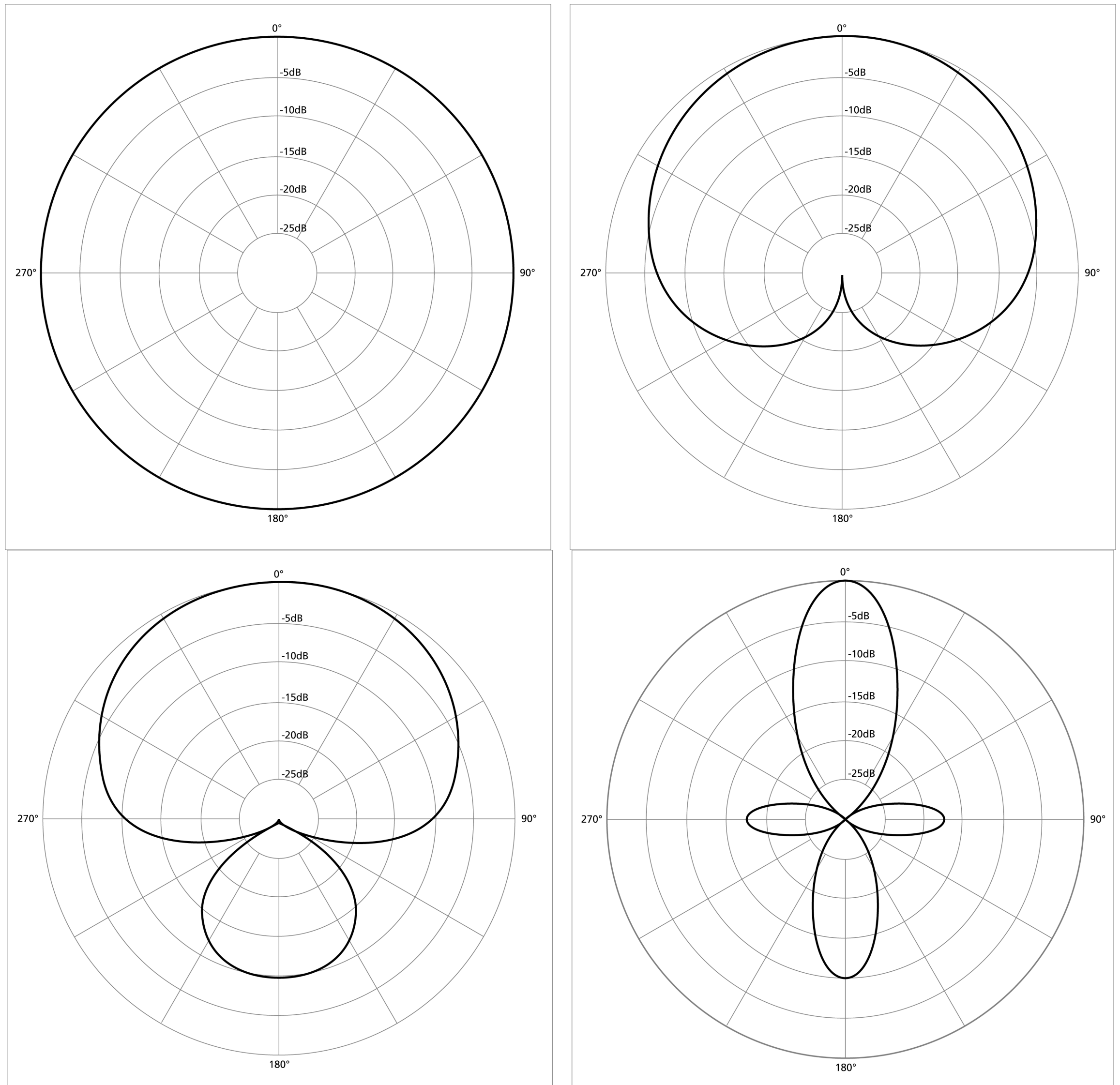
Difference Farfield/Nearfield

- Nearfield: More energy, less distortion, captures sound source well.

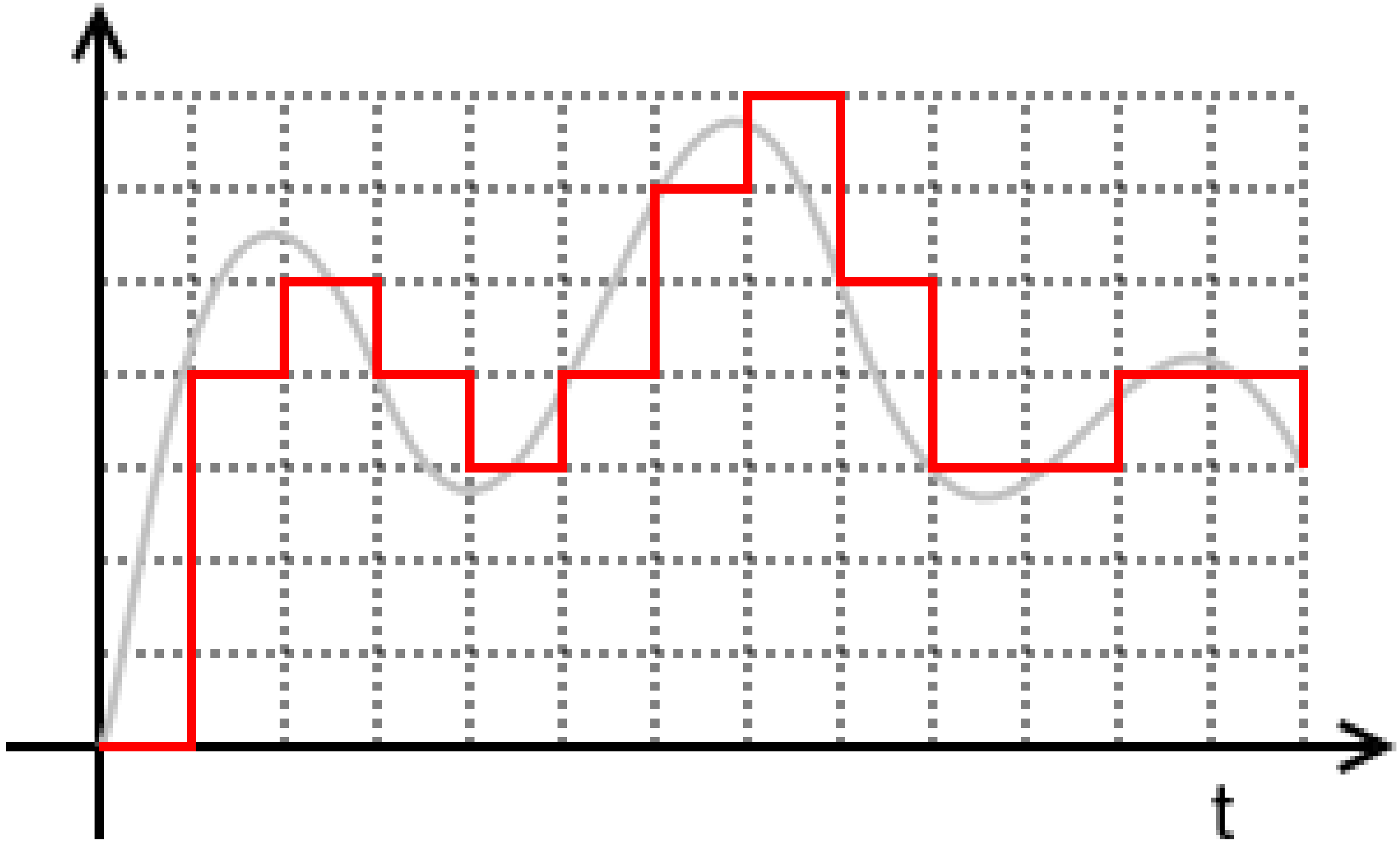
- Farfield: Captures environment with sound source, “better for forensics”, processing often slower.

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

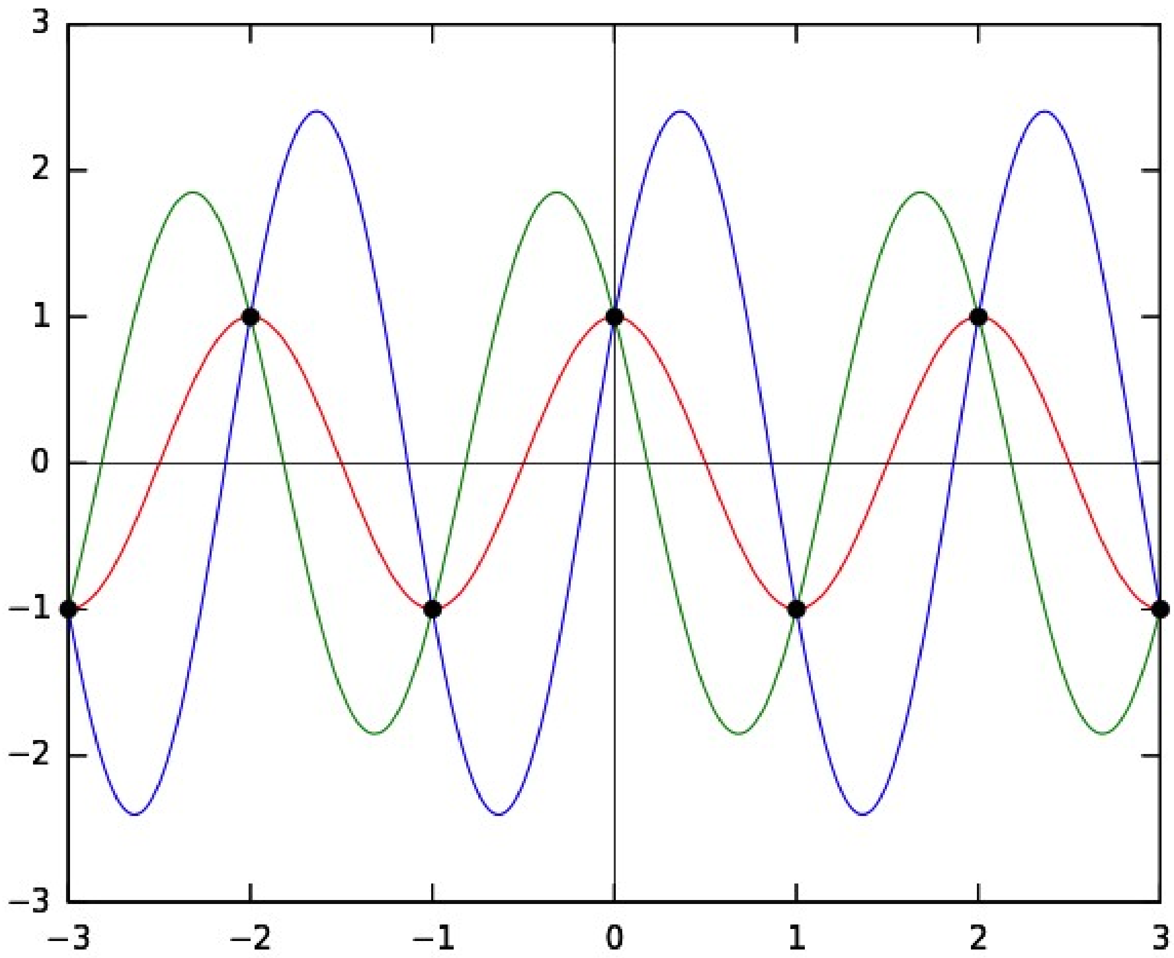
Microphone Directionality



Digitization of Sound



Remember: Nyquist Limit!

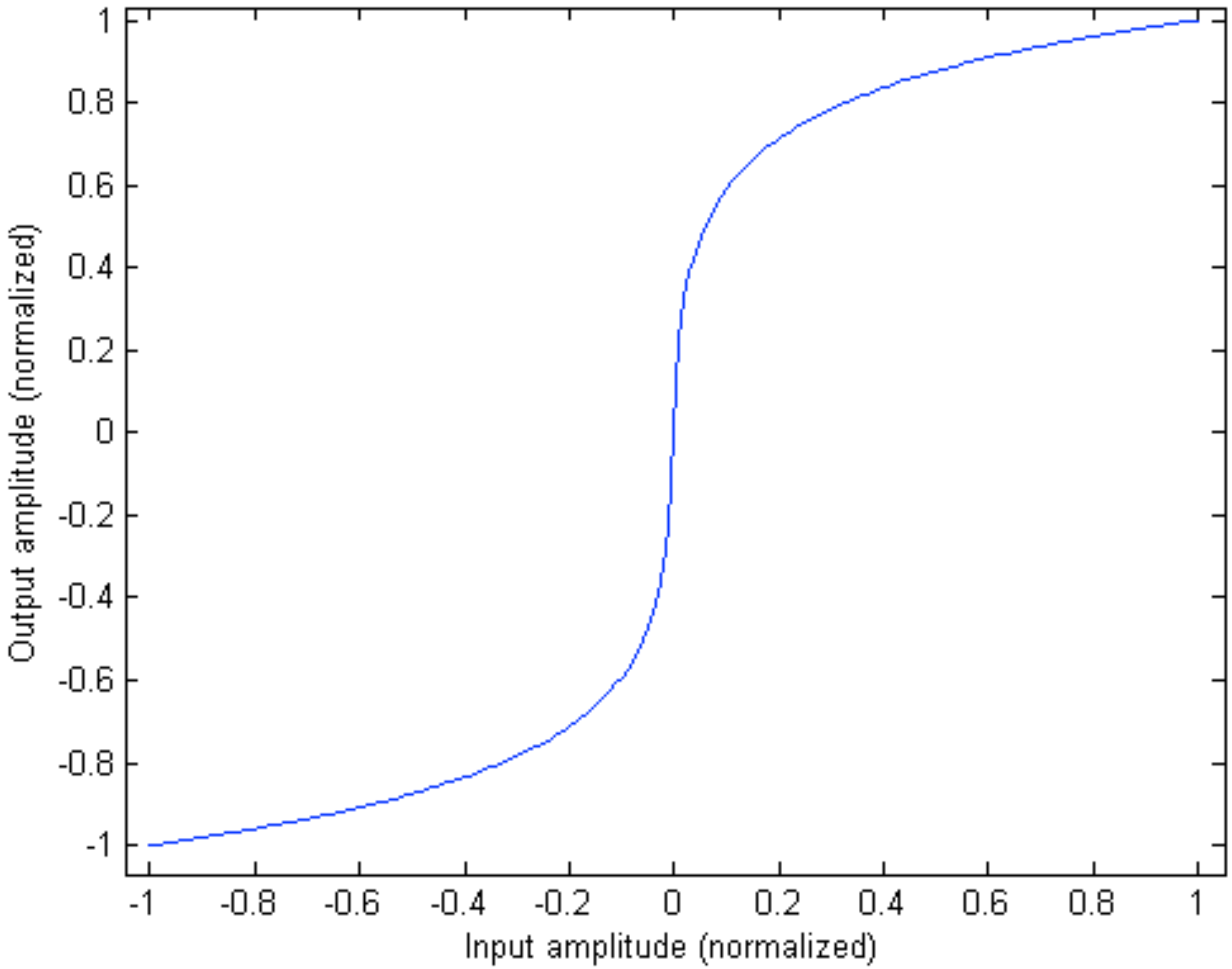


Math: See Draft Chapter 3 of Friedland & Jain on mm-creole.org

Common Recording Resolutions

- 8000Hz, 8-bit log. companded ~ 11 bit uncompanded (a/μ-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

μ-law Companding



Next Week

- Intro to Audio Analysis