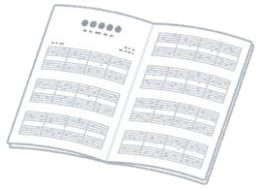
PAT 463/563 (Fall 2026)

Music & AI

Lecture 5: Digital Audio

Instructor: Hao-Wen (Herman) Dong

UNIVERSITY OF MICHIGAN

Music Representations

Symbolic music representations

- **Text-based**
  - MIDI

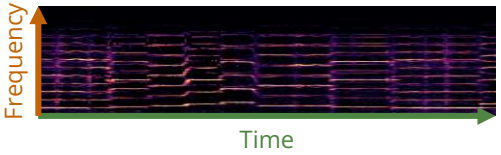
```
Program_change_0,
Note_on_60, Time_shift_2, Note_off_60,
Note_on_60, Time_shift_2, Note_off_60,
Note_on_76, Time_shift_2, Note_off_67,
Note_on_67, Time_shift_2, Note_off_67,
...
```

- **Image-based**
  - Piano roll

Audio-domain music representations

- **Time series-based**
  - Waveform

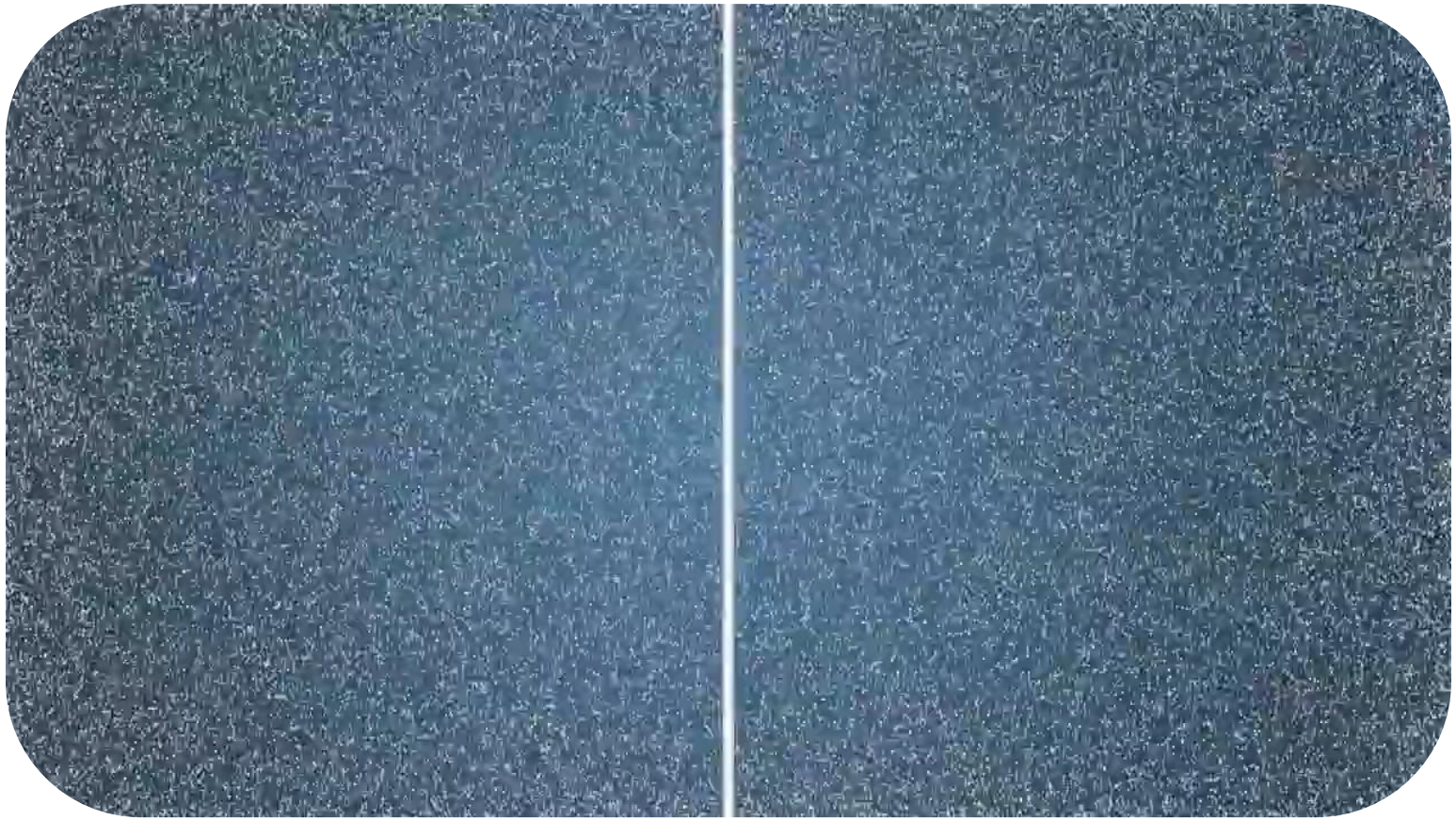
- **Image-based**
  - Spectrogram



Today's topic!

What is Sound?

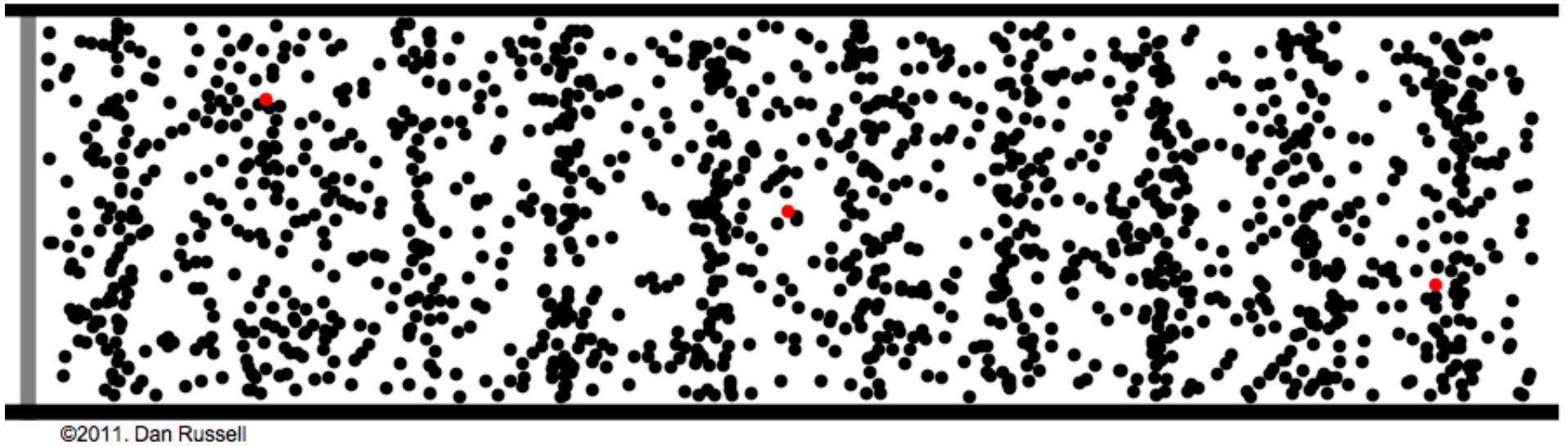
What is Sound?



youtu.be/aPswnDcteS4

jovansystem, "Sound Wave Animation," *YouTube*, Apr 12, 2012.

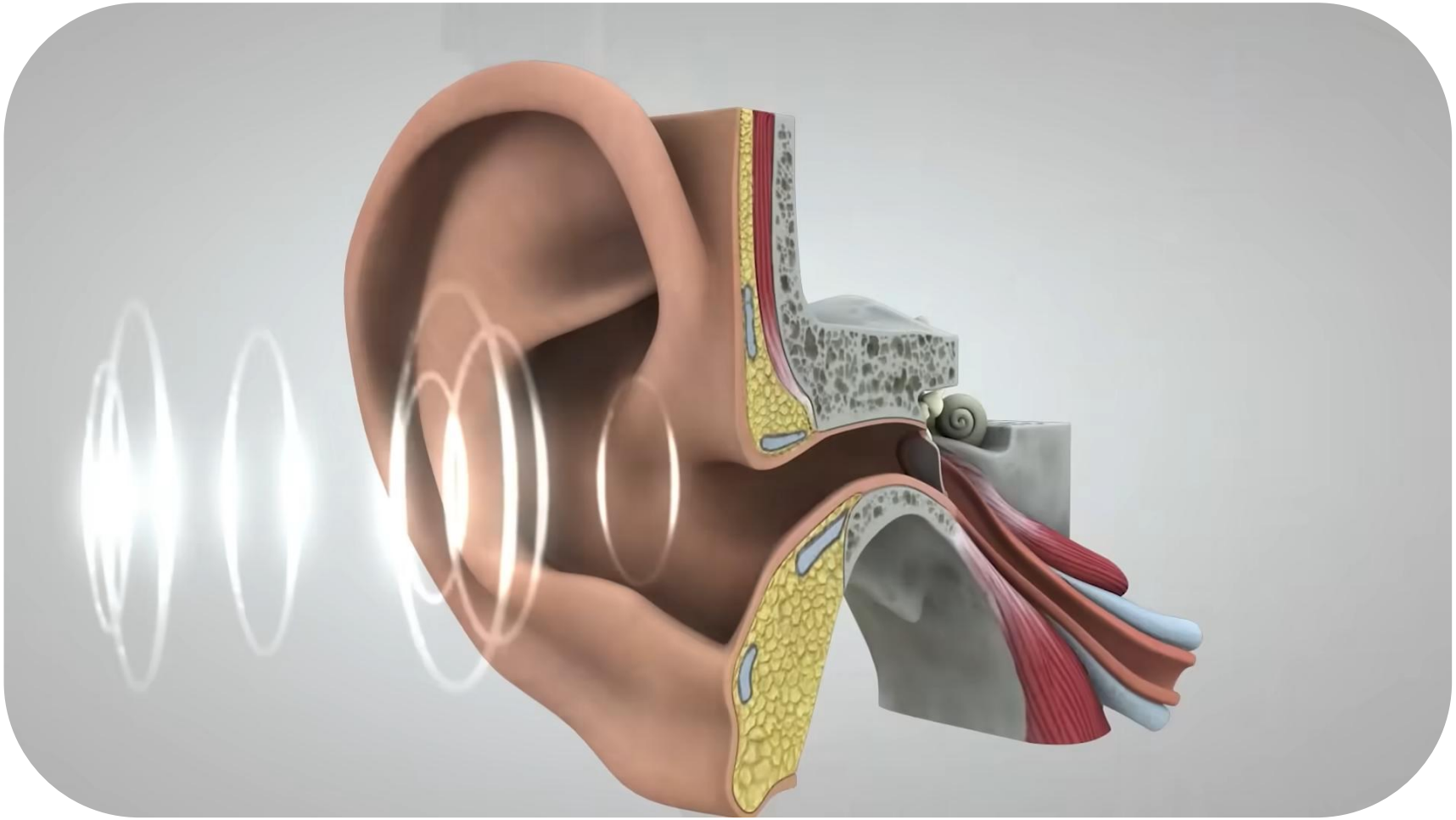
Sound is a Pressure Wave



(Source: Russell, 2011)

Dan Russell, "Longitudinal and Transverse Wave Motion," *Acoustics and Vibration Animations*, 1998 (updated on 2016).

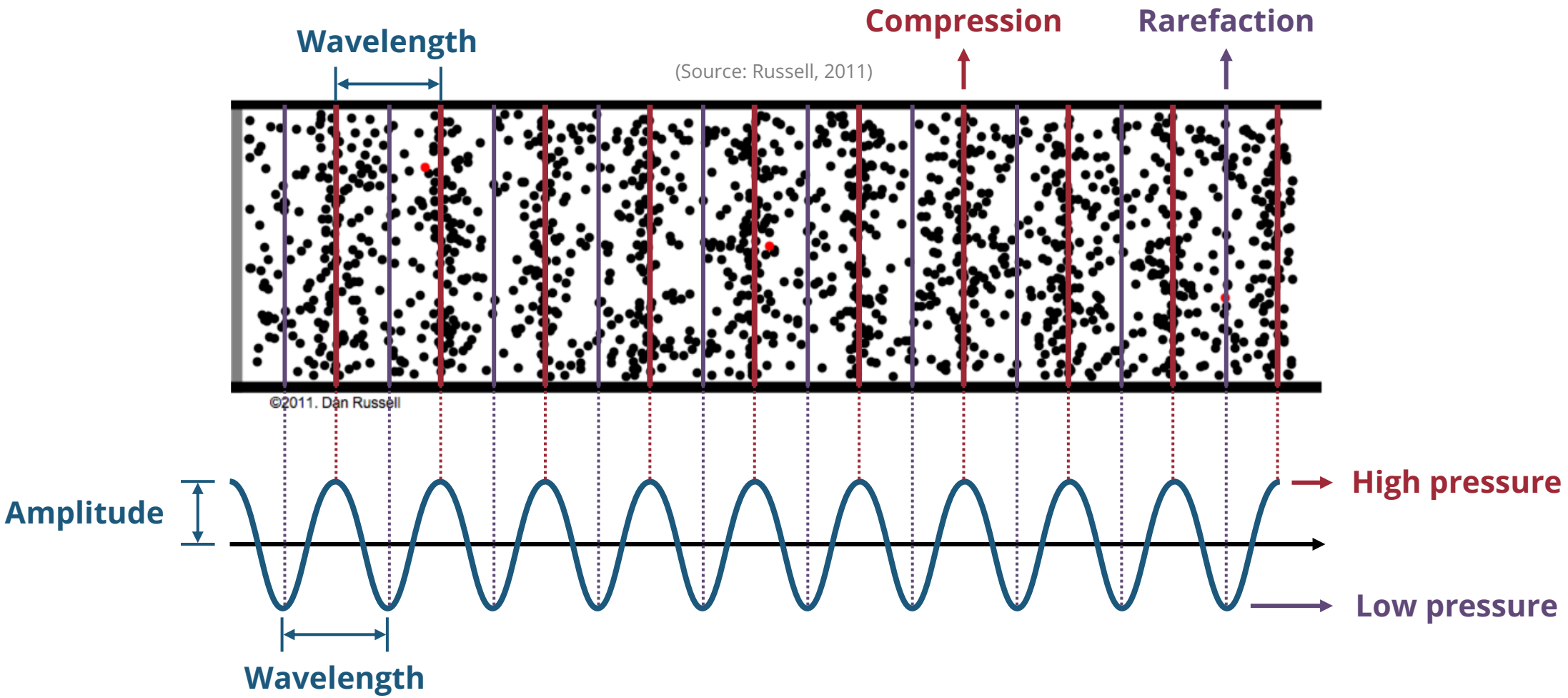
Human Auditory System



youtu.be/eQEaiZ2j9oc

National Institutes of Health (NIH), "Journey of Sound to the Brain," *YouTube*, November 21, 2018.

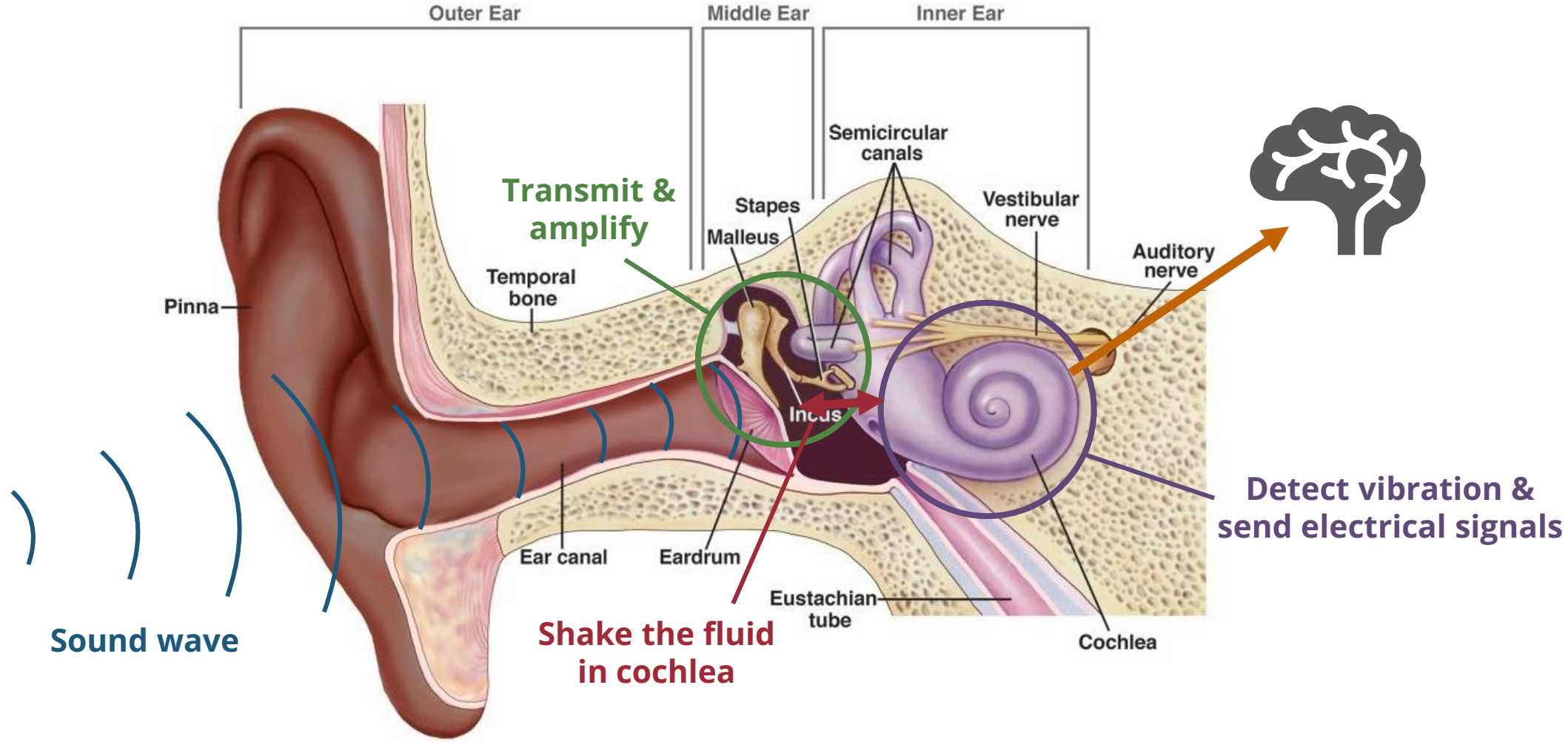
Longitudinal vs. Transverse Waves



Dan Russell, "Longitudinal and Transverse Wave Motion," *Acoustics and Vibration Animations*, 1998 (updated on 2016).

Auditory Perception

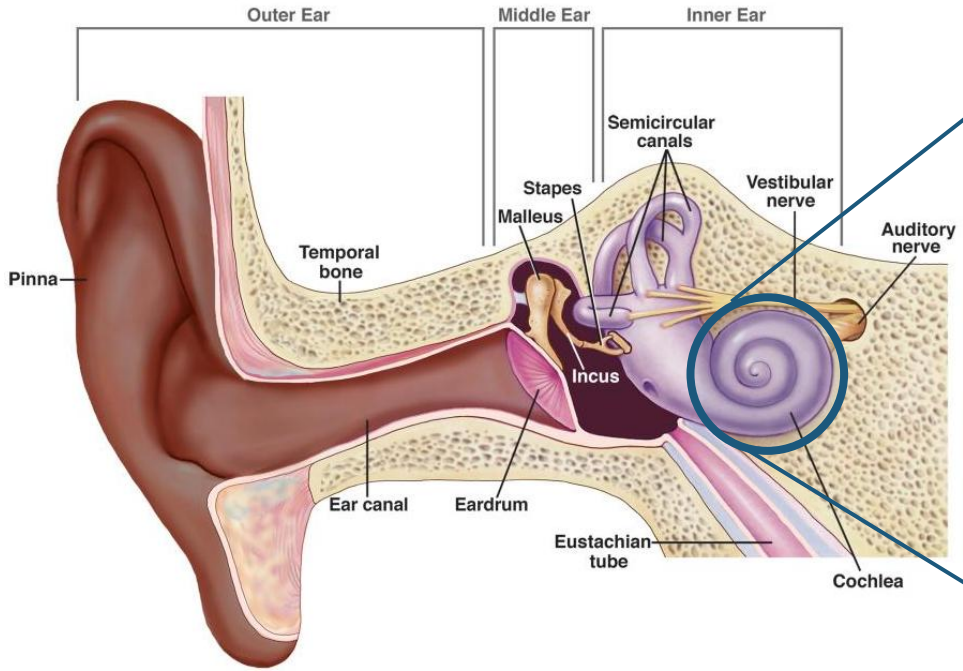
Human Ears



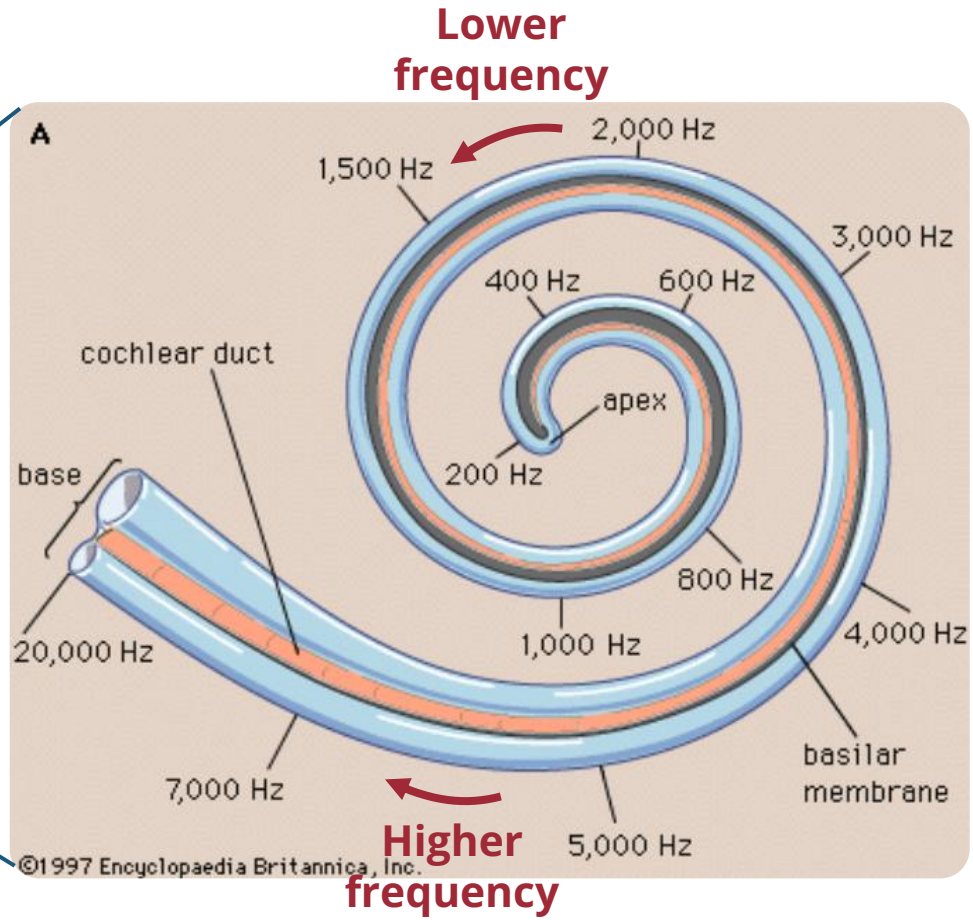
(Source: NIH/NIDCD)

National Institute on Deafness and Other Communication Disorders (NIDCD), "How Do We Hear?," National Institutes of Health (NIH), 2015.

Cochlea in the Inner Ear



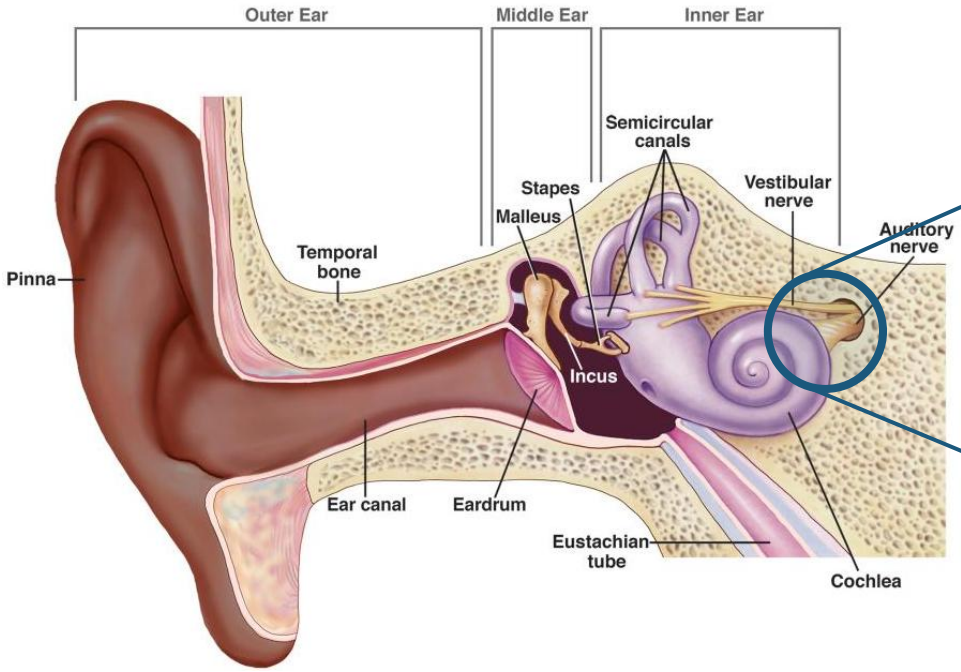
(Source: NIH/NIDCD)



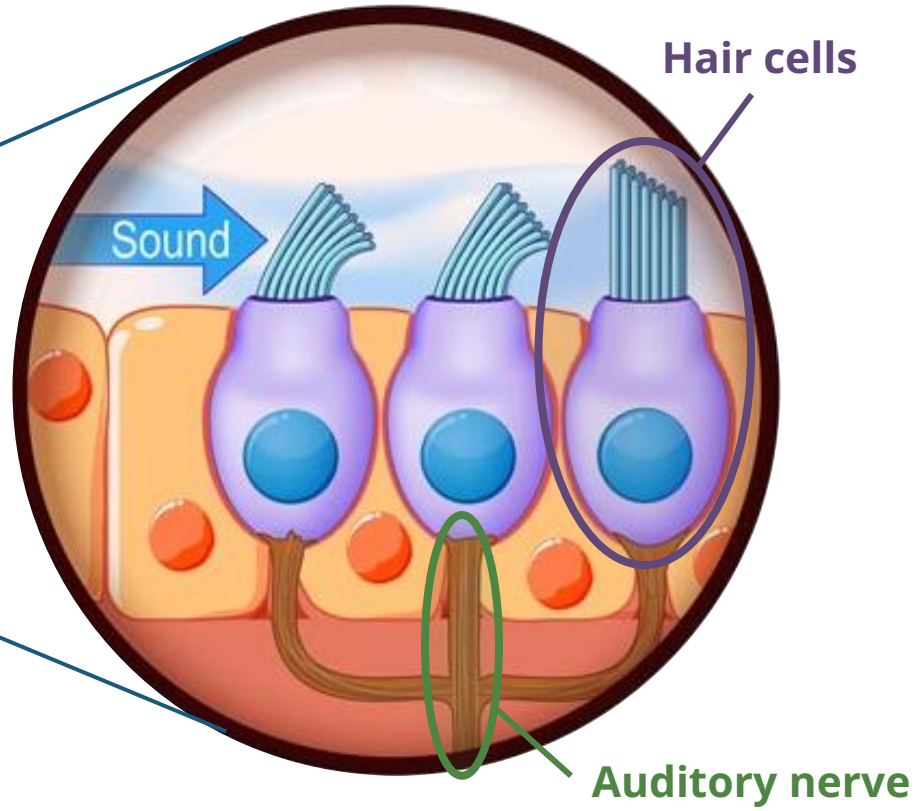
(Source: Britannica)

National Institute on Deafness and Other Communication Disorders (NIDCD), "How Do We Hear?," National Institutes of Health (NIH), 2015.
Joseph E. Hawkins, "Transmission of sound within the inner ear," B*ritannica*.

Hair Cells in the Cochlea



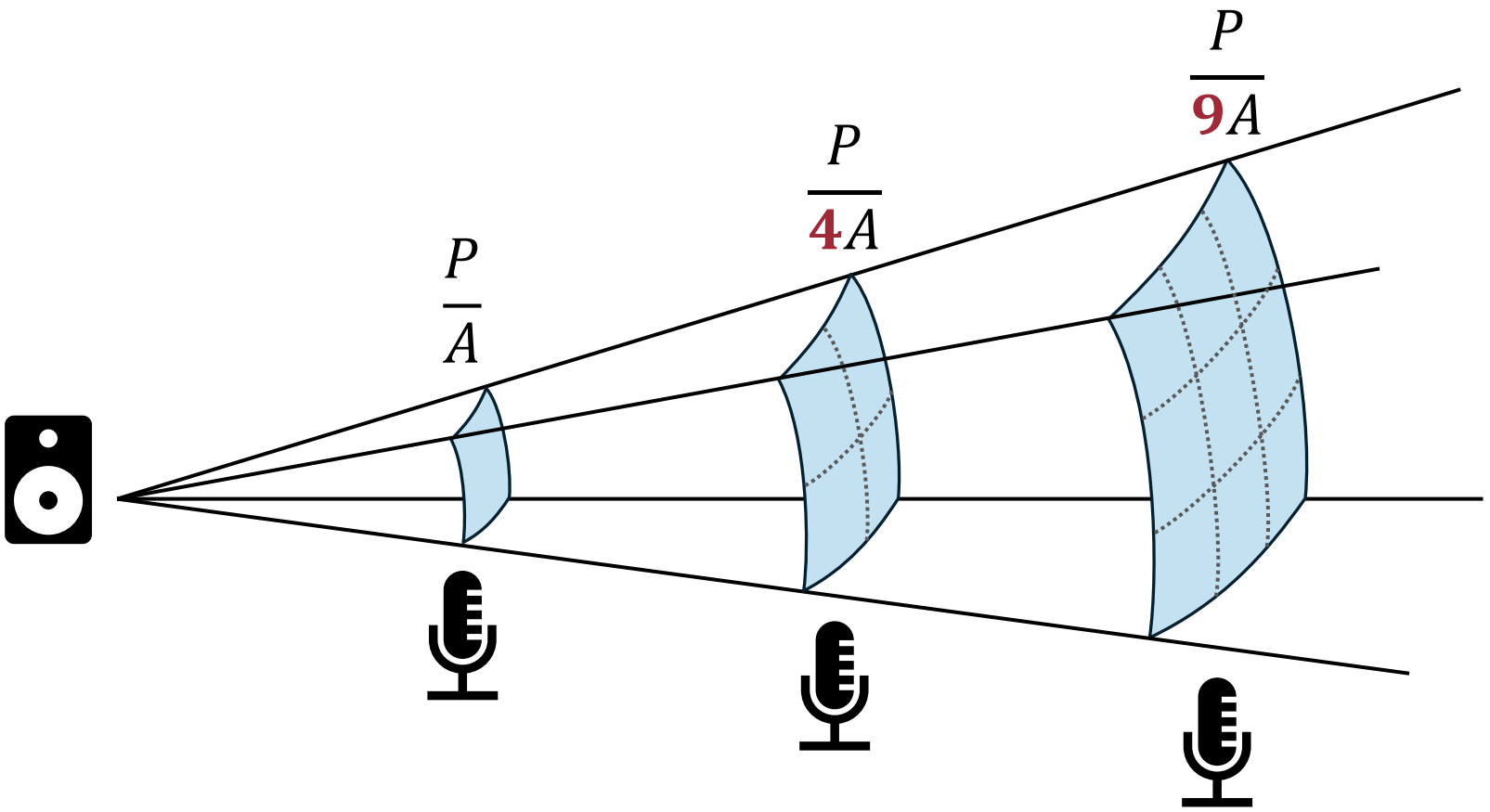
(Source: NIH/NIDCD)



(Source: COSMOS Magazine)

National Institute on Deafness and Other Communication Disorders (NIDCD), "How Do We Hear?," National Institutes of Health (NIH), 2015.
Matilda Handsley-Davis, "Genetic discovery may help scientists reverse hearing loss," *COSMOS Magazine*, May 5, 2022.

Sound Propagation & Inverse Square Law



Inverse Square Law

$$I \propto \frac{1}{r^2}$$

Sound Intensity & Decibels

- **Sound intensity** is defined as the sound power per unit area
  - Usually measured in **watt per square meter** (W/m^2)
- **Sound intensity *level*** is defined as

$$I_{\mathrm{dB}} \coloneqq 10 \log_{10} \left(\frac{I}{I_{REF}} \right)$$

  - $I_{REF} \coloneqq 10^{-12} W/m^2$ is the **threshold of hearing** (TOH)
  - TOH: minimum sound intensity of a pure tone that a human can hear

Loudness Measure: Decibels

	Decibels	Intensity	Type of sound	
😵	**130**	10	Artillery fire at close proximity (**threshold of pain**)	
	120	1	Amplified rock music; near jet engine	
	110	10^{-1}	Loud orchestral music, in audience	
😣	**100**	10^{-2}	Electric saw	
	90	10^{-3}	Bus or truck interior	
😐	**80**	10^{-4}	Automobile interior	
	70	10^{-5}	Average street noise; loud telephone bell	
🙂	**60**	10^{-6}	**Normal conversation**; business office	
	50	10^{-7}	Restaurant; private office	
	40	10^{-8}	Quiet room in home	
😌	**30**	10^{-9}	Quiet lecture hall; bedroom	
	20	10^{-10}	Radio, television, or recording studio	
	10	10^{-11}	Soundproof room	
😮‍💨	**0**	10^{-12}	Absolute silence (**threshold of hearing**)	

(Unit: W/m^2) (Source: Britannica)

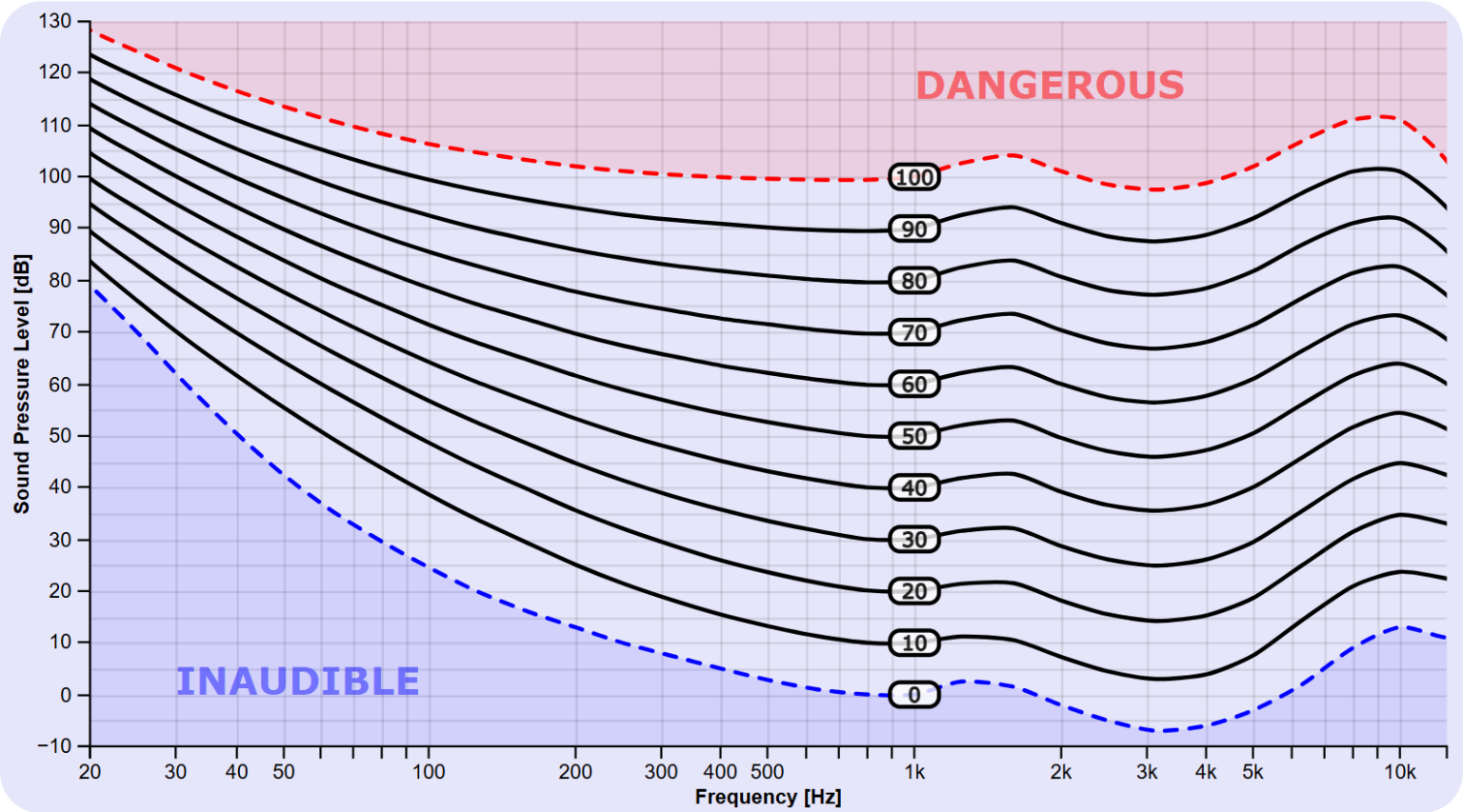
Richard E. Berg, "The decibel scale," *Britannica*.

🤔 Common Gains in Decibels

- **+10 dB = ?x intensity**
  - **10x intensity** (3.16x amplitude)
- **+3 dB ≈ ?x intensity**
  - **2x intensity** (1.414x amplitude)

- **+20 dB = ?x amplitude**
  - **10x amplitude** (100x intensity)
- **+6 dB ≈ ?x amplitude**
  - **2x amplitude** (4x intensity)

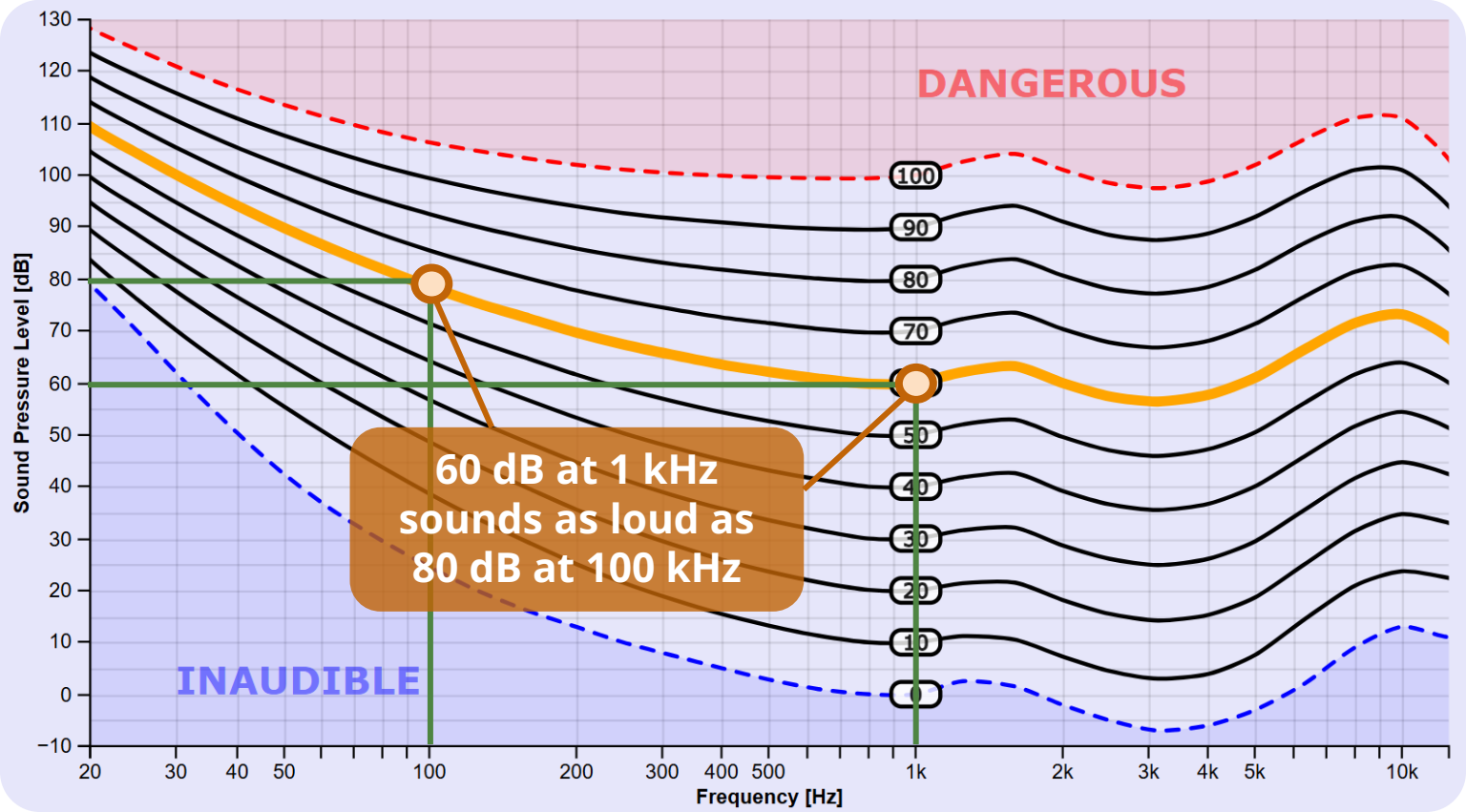
Loudness Perception: Equal-loudness Contours



(Source: Parker, 2024)

Noah J. Parker, "Normal Equal-Loudness-Level Contours," *Math of Sound*, 2024.

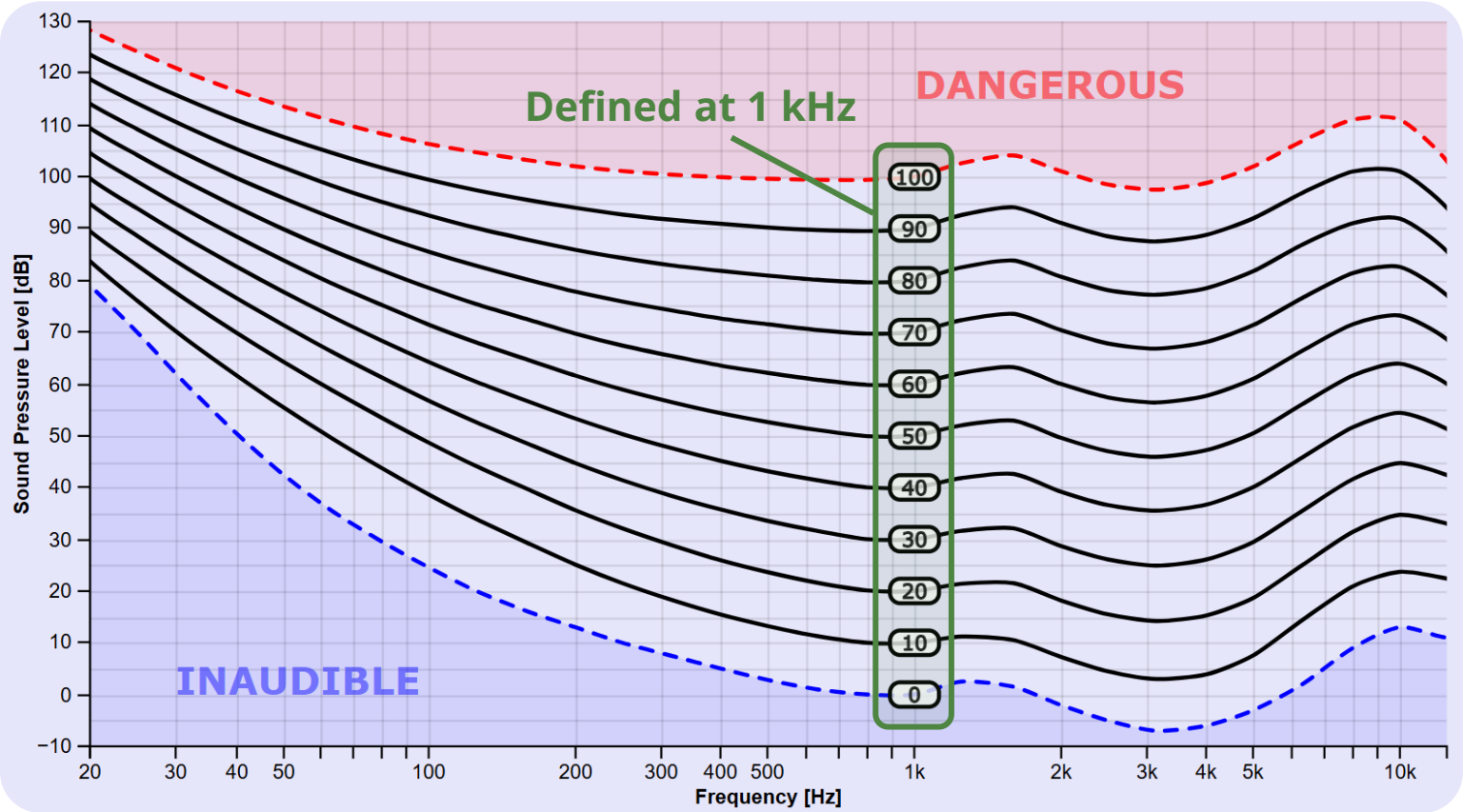
Loudness Perception: Equal-loudness Contours



(Source: Parker, 2024)

Noah J. Parker, "Normal Equal-Loudness-Level Contours," *Math of Sound*, 2024.

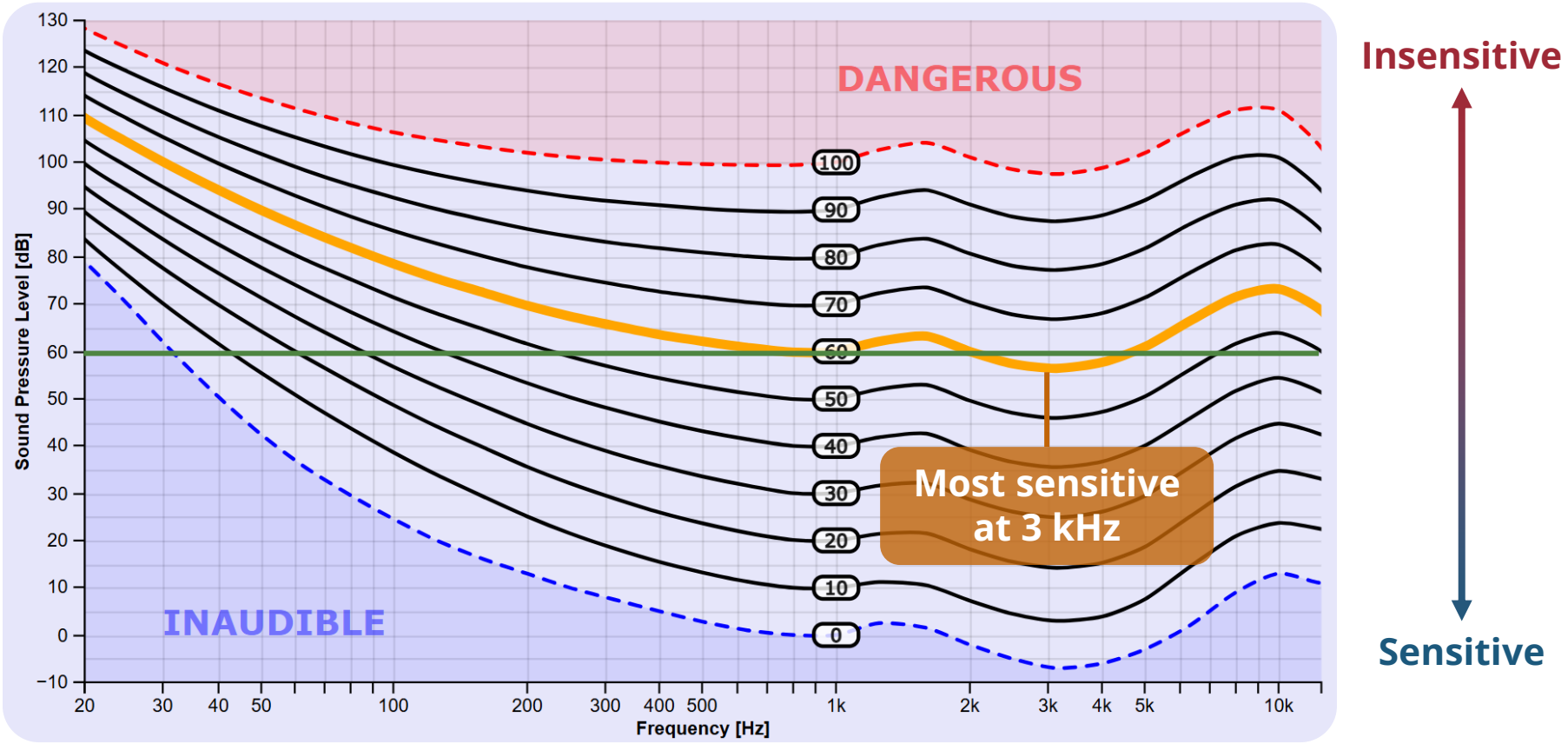
Loudness Perception: Equal-loudness Contours



(Source: Parker, 2024)

Noah J. Parker, "Normal Equal-Loudness-Level Contours," *Math of Sound*, 2024.

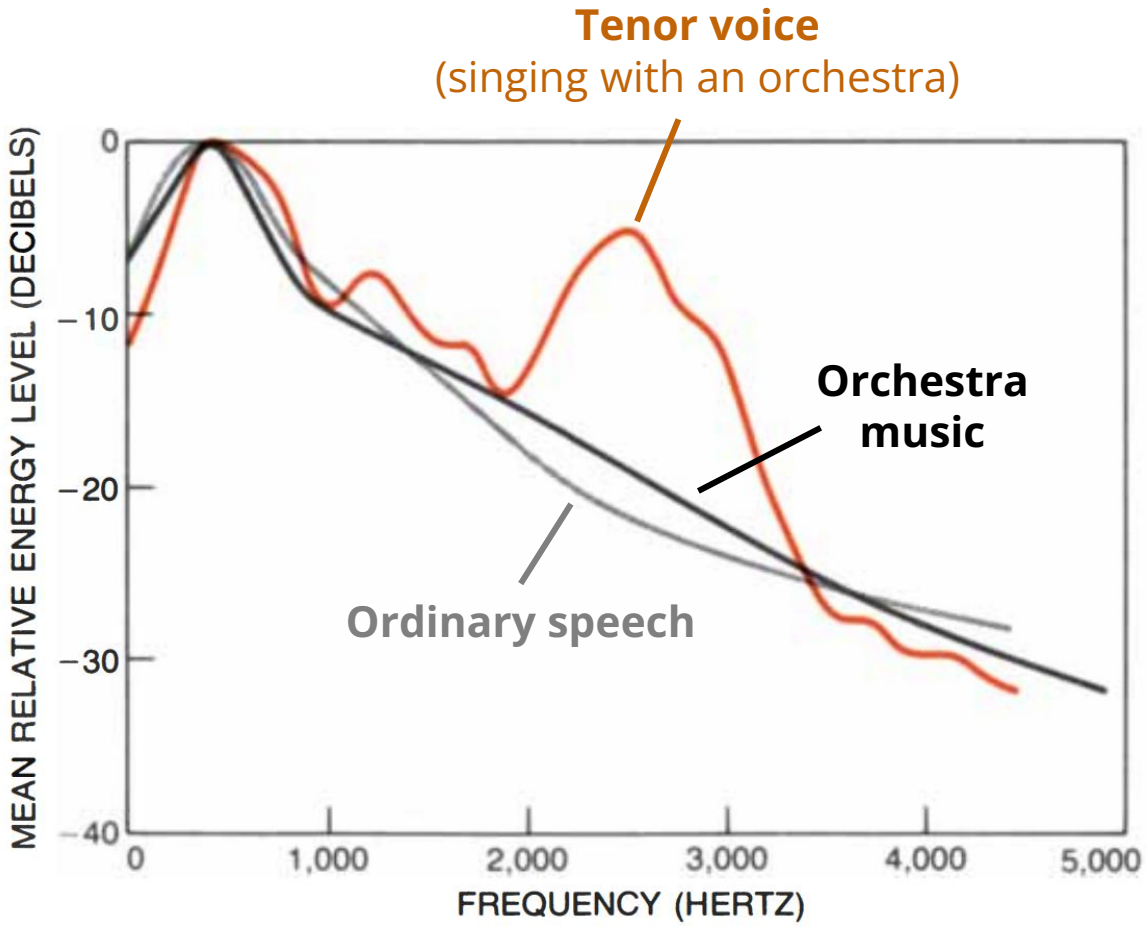
Loudness Perception: Equal-loudness Contours



(Source: Parker, 2024)

Noah J. Parker, "Normal Equal-Loudness-Level Contours," *Math of Sound*, 2024.

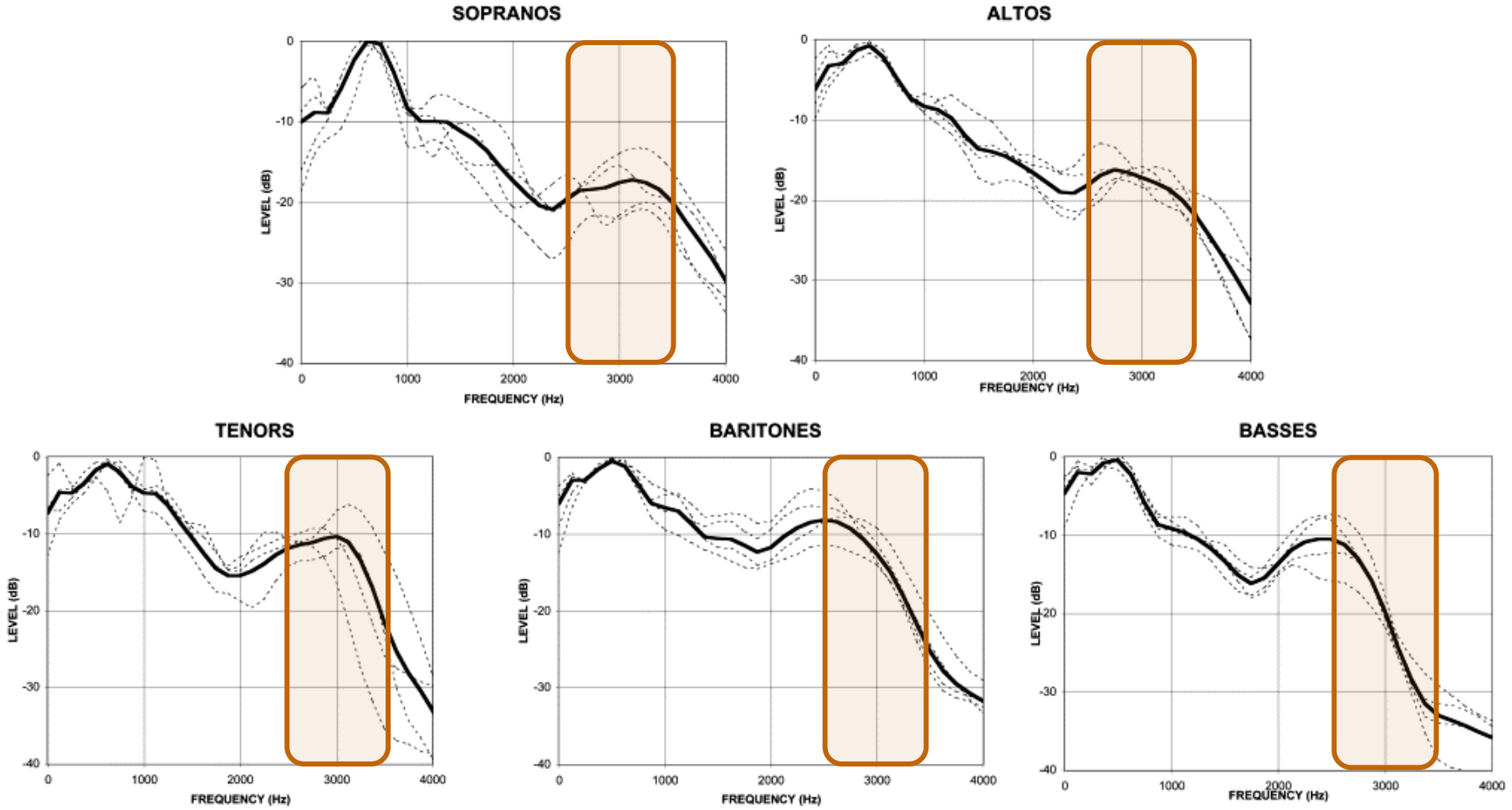
Singer's Formants (Sundberg, 1991)



(Source: Sundberg, 1977)

Johan Sundberg, "The Acoustics of the Singing Voice," *Scientific American*, 236(3):82-91, 1977.

Singer's Formants (Sundberg, 1991)

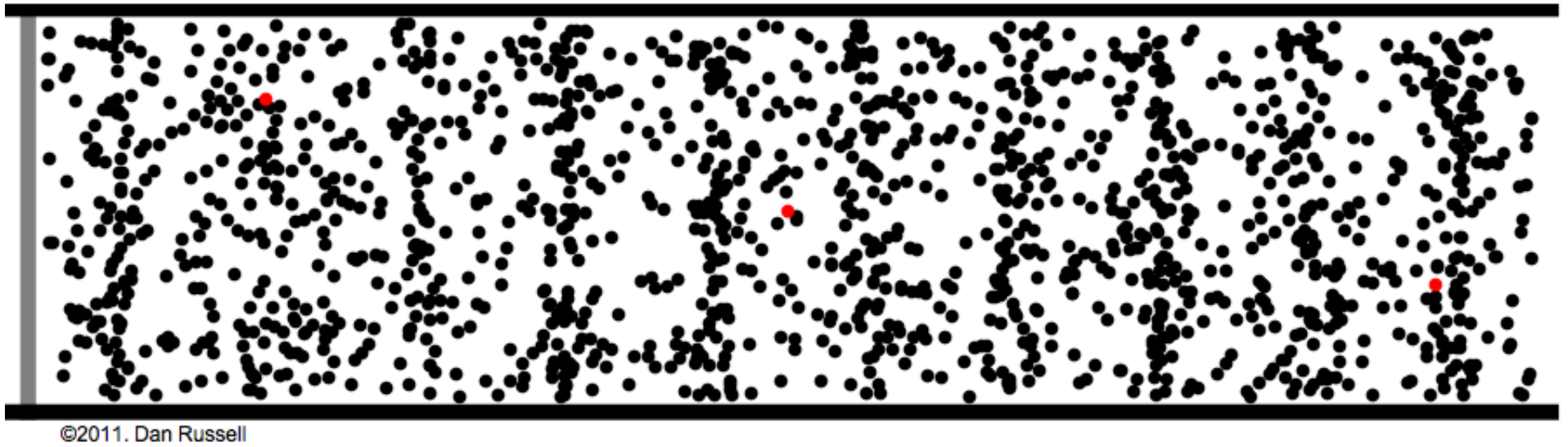


(Source: Sundberg, 1977)

Johan Sundberg, "The Acoustics of the Singing Voice," *Scientific American*, 236(3):82-91, 1977.

Analog Audio

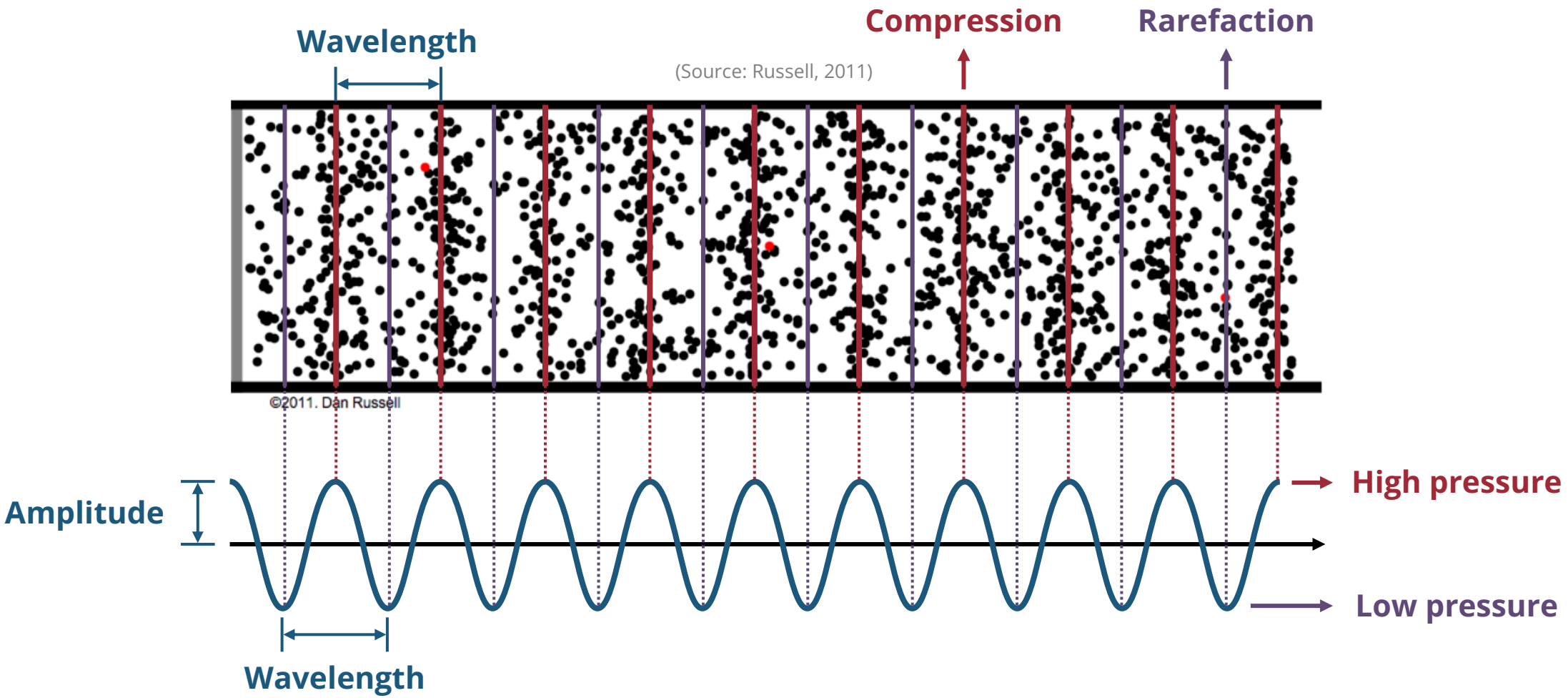
Sound is a Pressure Wave



(Source: Russell, 2011)

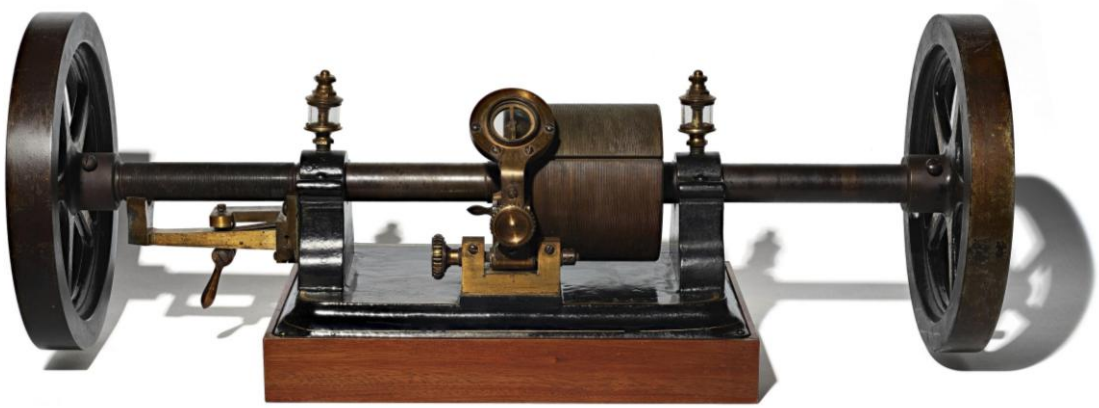
Dan Russell, "Longitudinal and Transverse Wave Motion," *Acoustics and Vibration Animations*, 1998 (updated on 2016).

Longitudinal vs. Transverse Waves

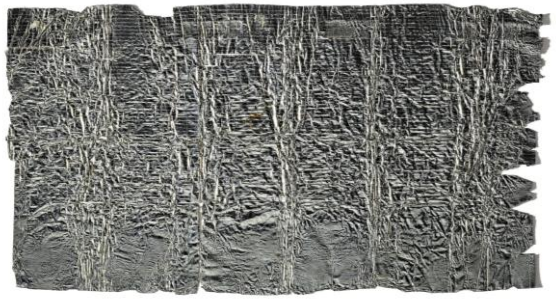


Dan Russell, "Longitudinal and Transverse Wave Motion," *Acoustics and Vibration Animations*, 1998 (updated on 2016).

Edison's Tinfoil Phonograph (1877)



(Source: National Museum of American History)



(Source: National Museum of American History)

Edison tin foil recording (fragment)

(Source: Library of Congress)

americanhistory.si.edu/explore/exhibitions/americas-listening/online/edisons-talking-machine

📖 Reading: Recording on Tin Foils

youtu.be/KRnTFiXkNFw

Edison Home Phonograph with Wax Cylinders (1896)

(Source: National Museum of American History)

(Source: M. Dupres)

americanhistory.si.edu/explore/exhibitions/americas-listening/online/edisons-talking-machine
M. Dupres, CC BY-SA 4.0, via Wikimedia Commons

Edison Home Phonograph at Stearns Collection

STEARNS COLLECTION OF MUSICAL INSTRUMENTS

Photos taken on April 3, 2025

Edison Home Phonograph with Wax Cylinders (1896)

(Source: Holger.Ellgaard)

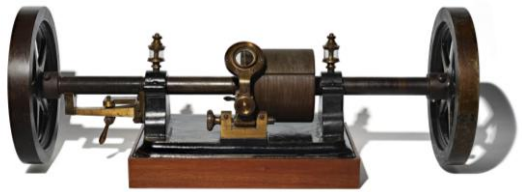
Holger.Ellgaard, CC BY-SA 3.0, via Wikimedia Commons

Recording on Wax Cylinders

youtu.be/1jeSpGfzlY8

Evolution of Phonograph

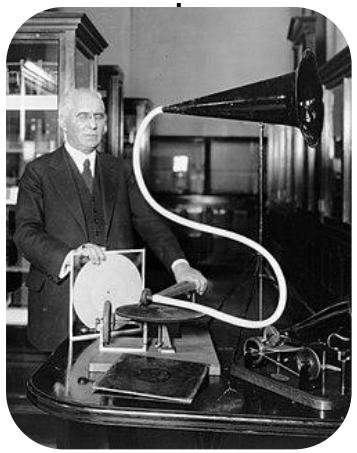
(Source: National Museum of American History)



1877

(Source: National Museum of American History)

1892



1896

(Source: Norman Bruderhofer)

1907

(Source: FredrikT)

1930s

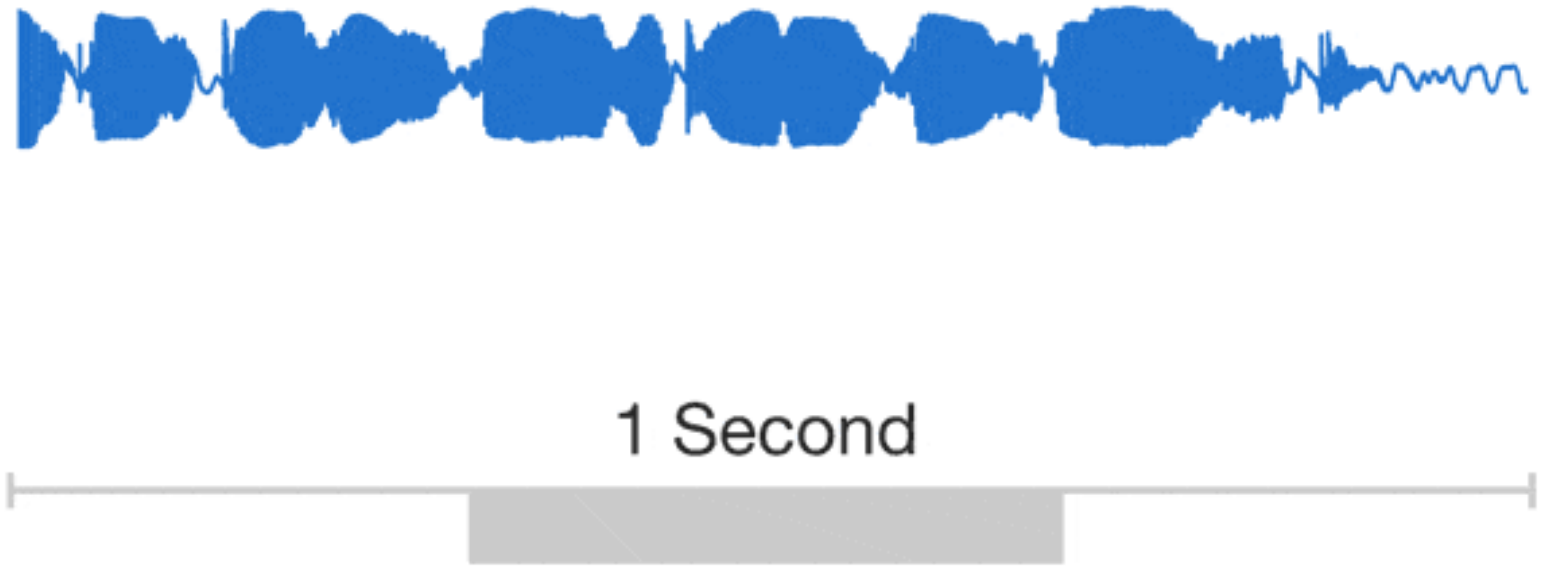
(Source: 32bitmaschine)

1972

americanhistory.si.edu/explore/exhibitions/americas-listening/online/edisons-talking-machine
Norman Bruderhofer, CC BY-SA 2.5, via Wikimedia Commons
FredrikT, CC BY-SA 3.0, via Wikimedia Commons
32bitmaschine, CC BY-SA 2.5, via Wikimedia Commons

Digital Audio

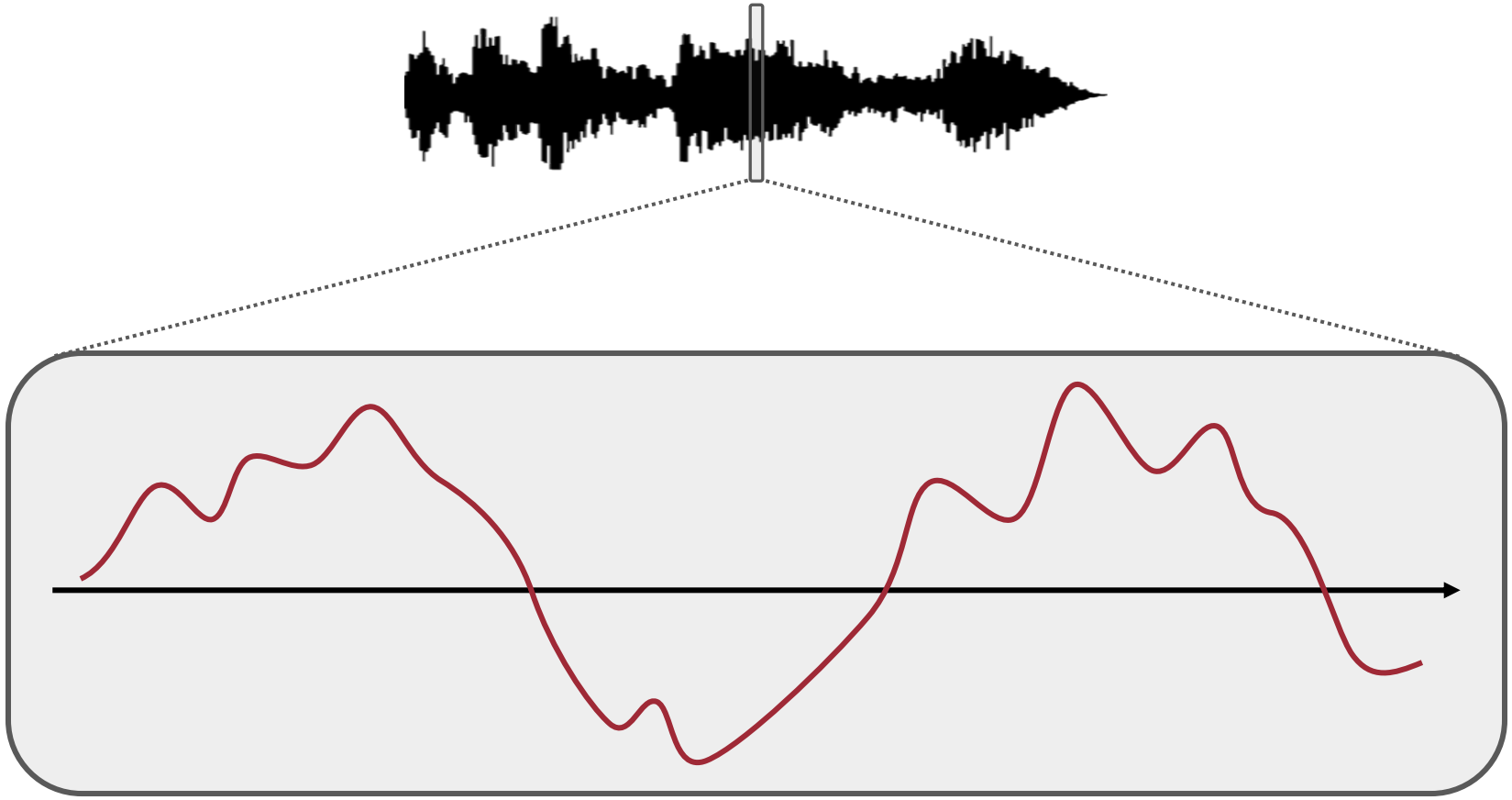
Digital Audio



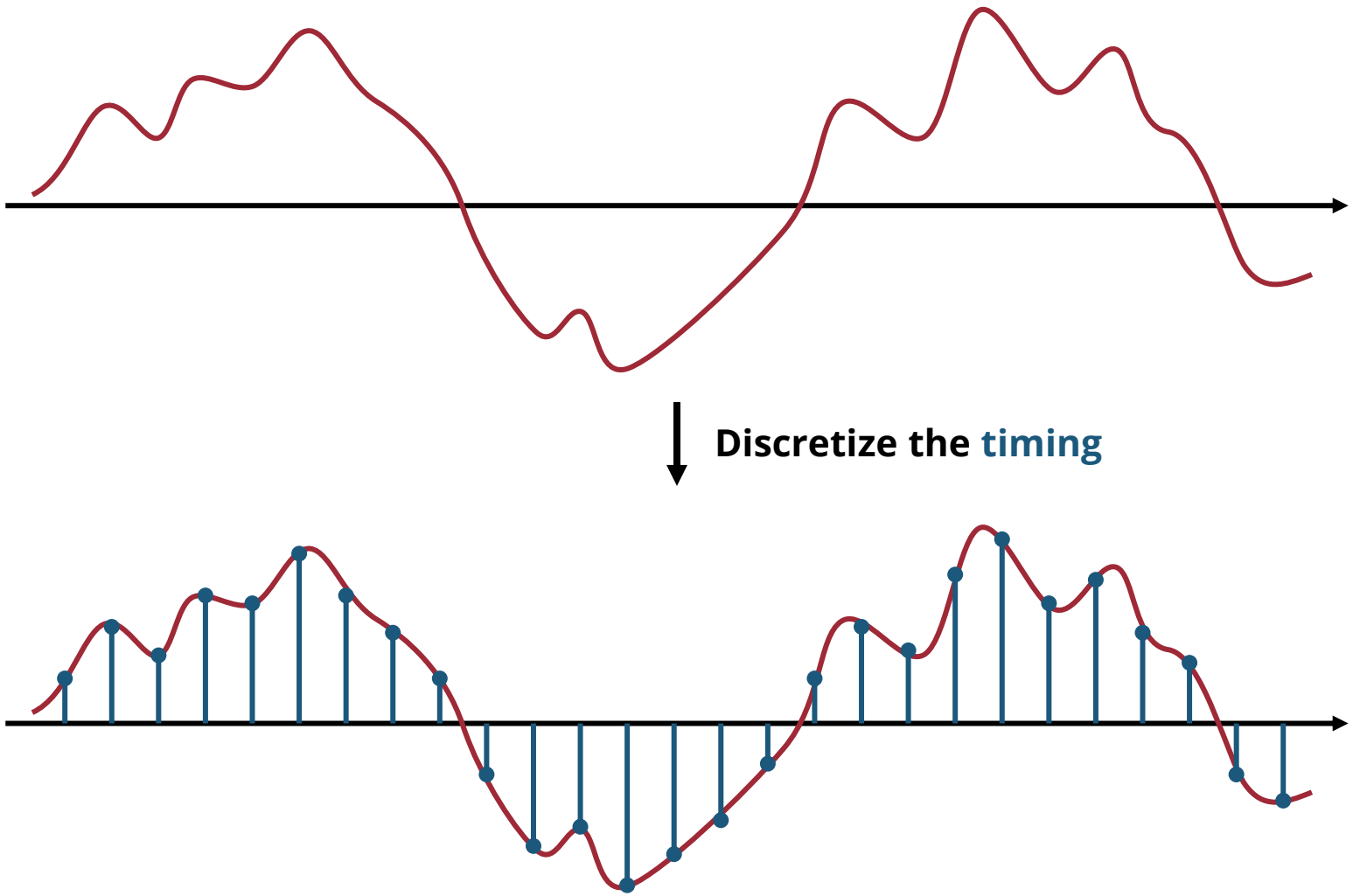
(Source: van den Oord et al., 2016)

Aaron van den Oord and Sander Dieleman, "WaveNet: A generative model for raw audio," *DeepMind Blog*, September 8, 2016.

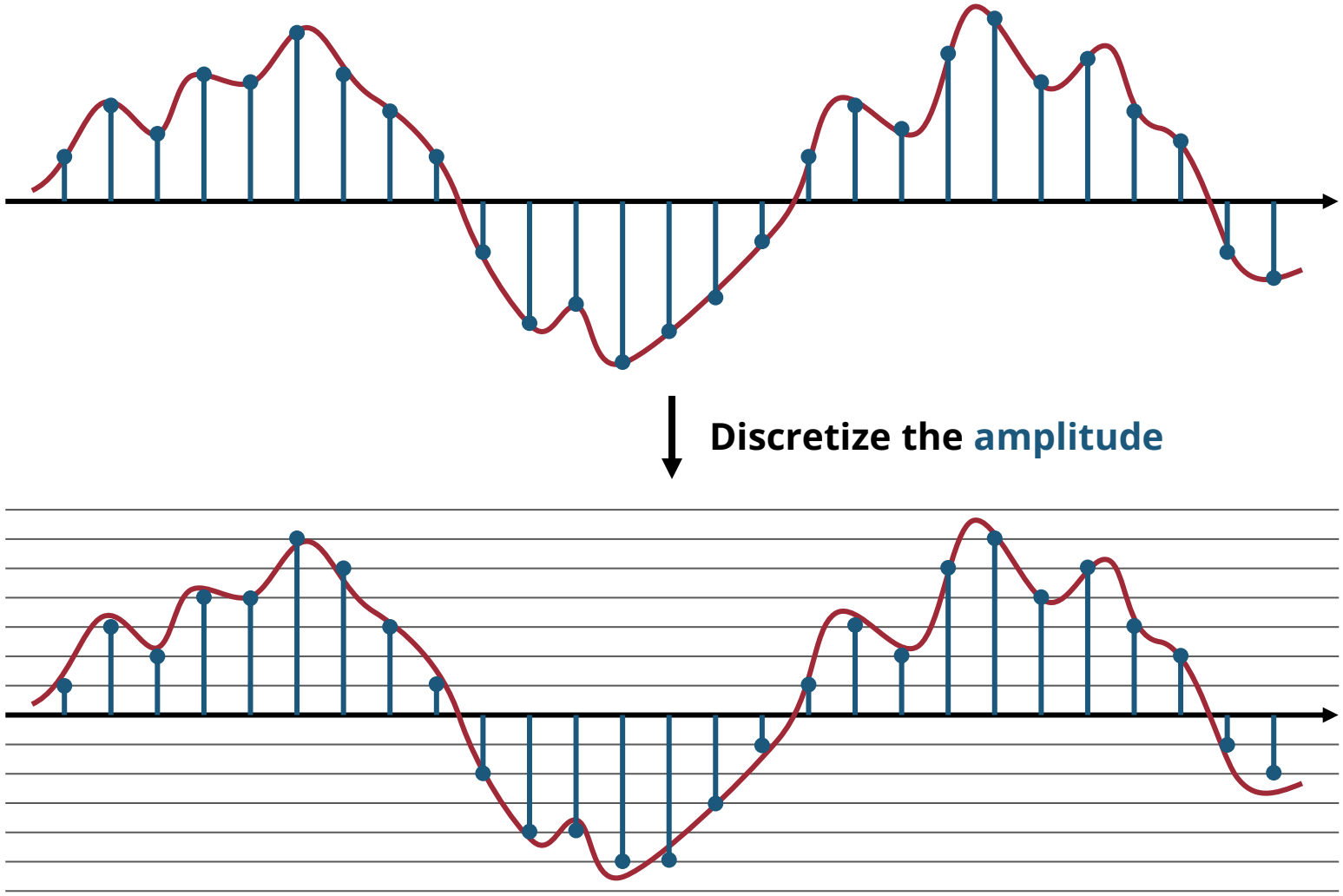
Digital Audio



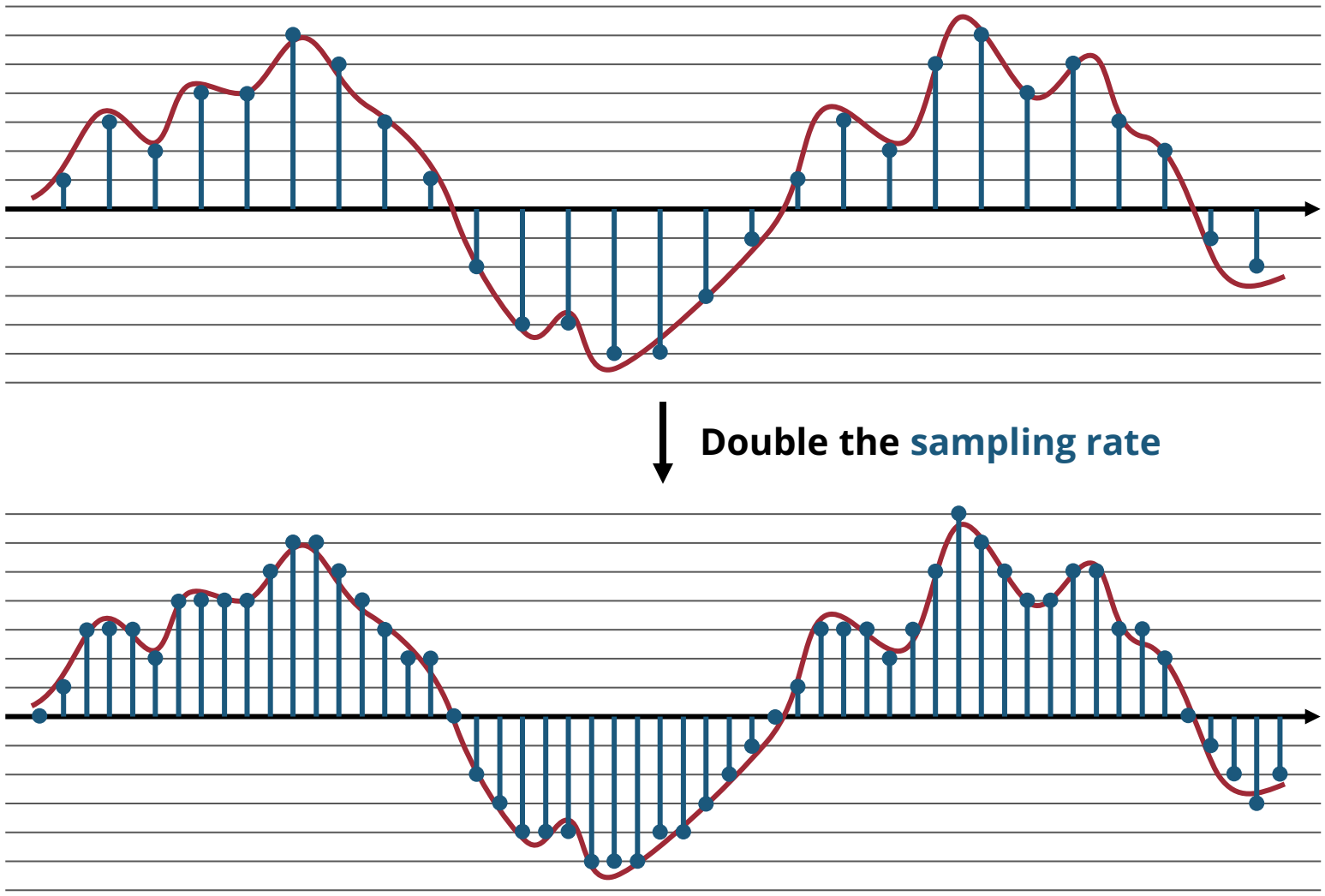
Digital Audio: Timing



Digital Audio: Amplitude



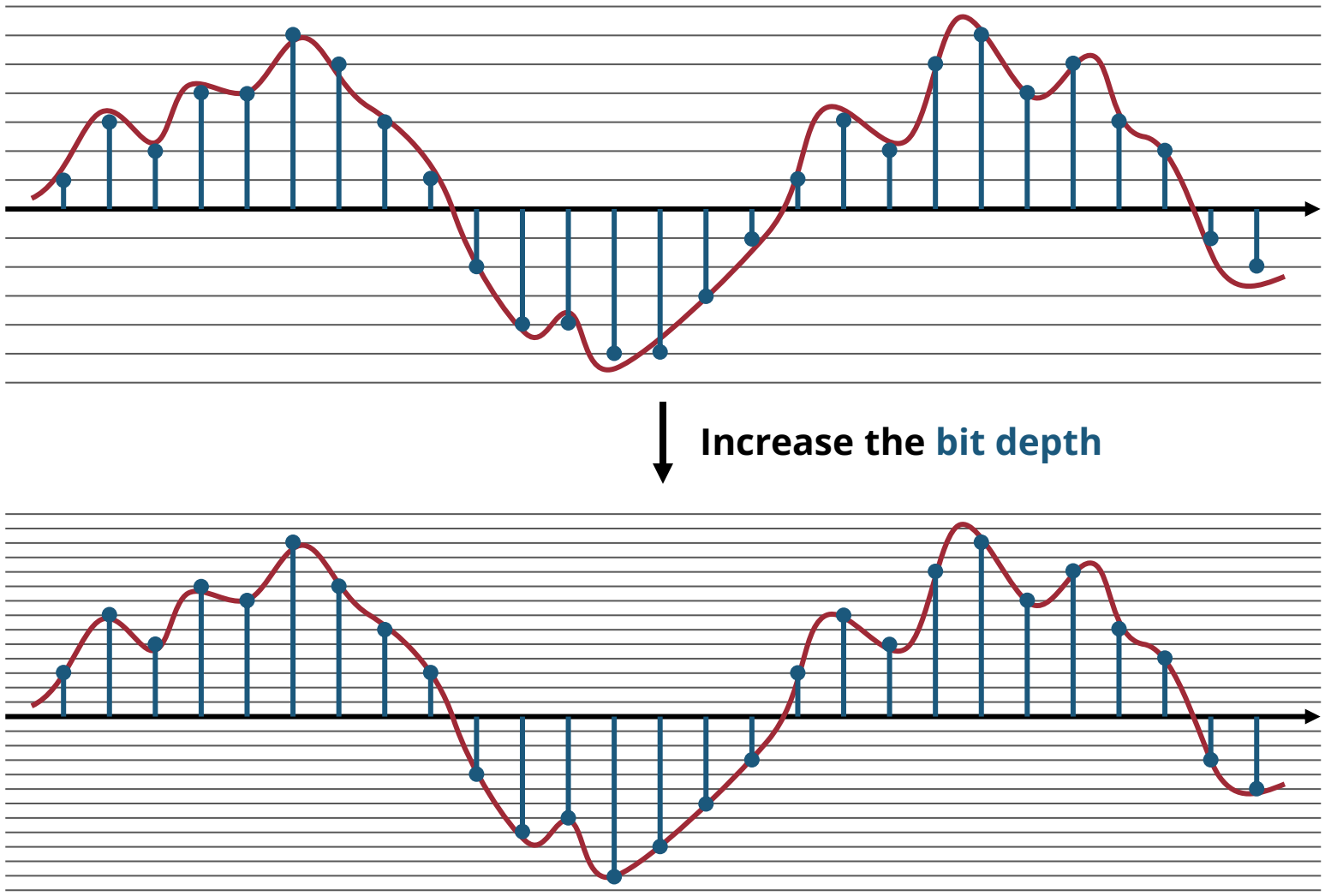
Sampling Rate: Temporal Resolution



Sampling Rate: Temporal Resolution

- **Definition**: **Number of samples per second**
  - How many times the “sound pressure” is measured per second
  - The higher the sampling rate, the lower the timing distortion
- **Common sampling rates**
  - **Telephone**: 8 kHz
  - **CD**: 44.1 kHz
  - **DVD**: 48 kHz
  - **Modern audio interfaces & DAWs**: 96 kHz, 192 kHz

Bit Depth: Dynamic Resolution



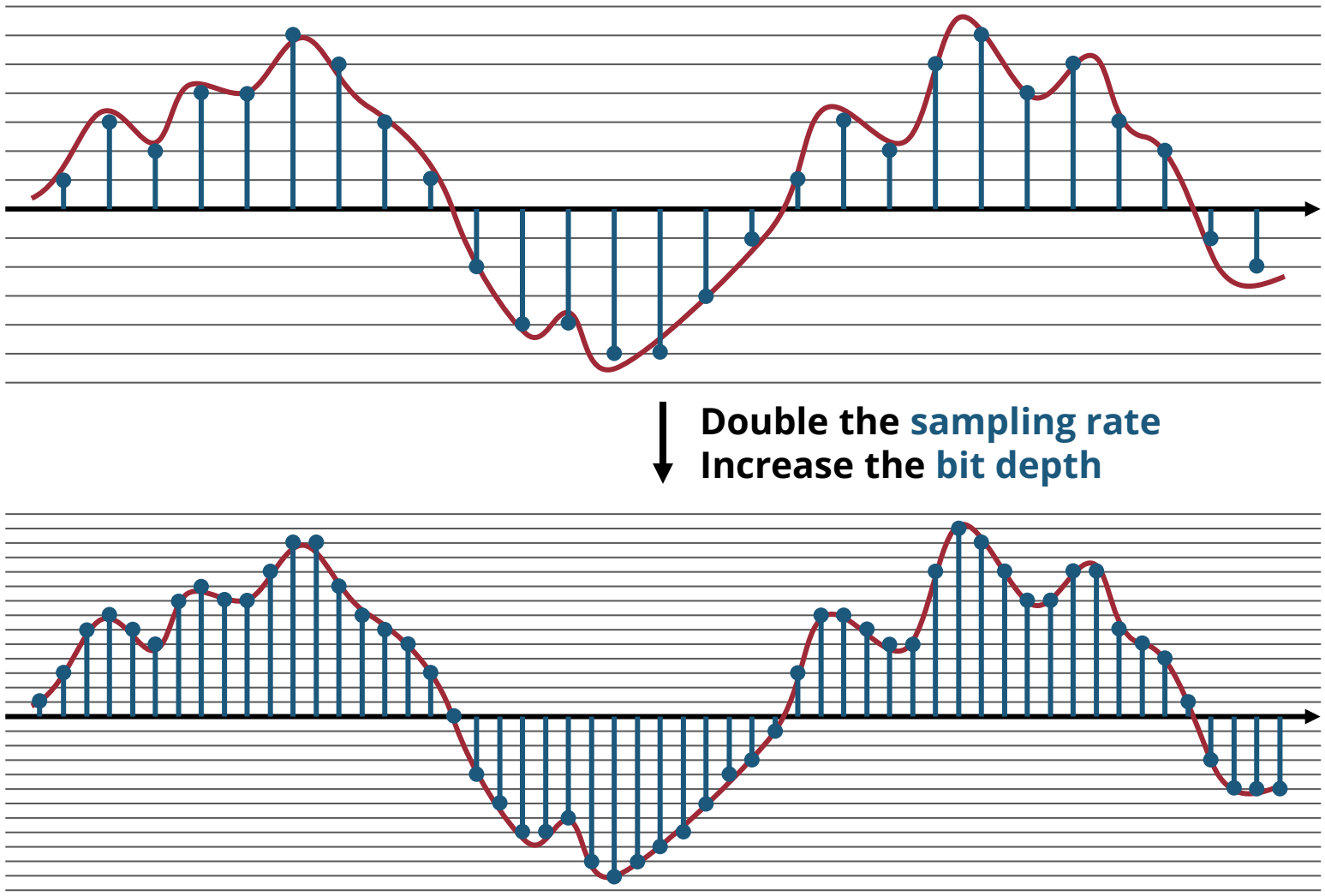
Bit Depth: Dynamic Resolution

- **Definition**: **Number of bits used to store each sample**
  - How many bits used to store the amplitude
  - The higher the bit depth, the lower the amplitude distortion
- **Common bit depth**
  - **Chiptunes**: 8 bit
  - **CD**: 16 bit
  - **Modern audio interfaces & DAWs**: 24 bit, 32 bit

Bit Depth

- **8 bit**: -128 to 127
- **16 bit**: -32,768 to 32,767
- **24 bit**: -8,388,608 to 8,388,607
- **32 bit**: 32-bit floating numbers

Resolution: Sampling Rate & Bit Depth



Sampling Rate vs. Bit Depth

- **Sampling rate** determines **how often** we measure the sound
- **Bit depth** determines **how precisely** we measure each point

Bit Depth vs. Bit Rate

- **Bit Depth**: **Number of bits used to store each sample**
  - Example: **CD quality** is **16bit/44.1kHz**

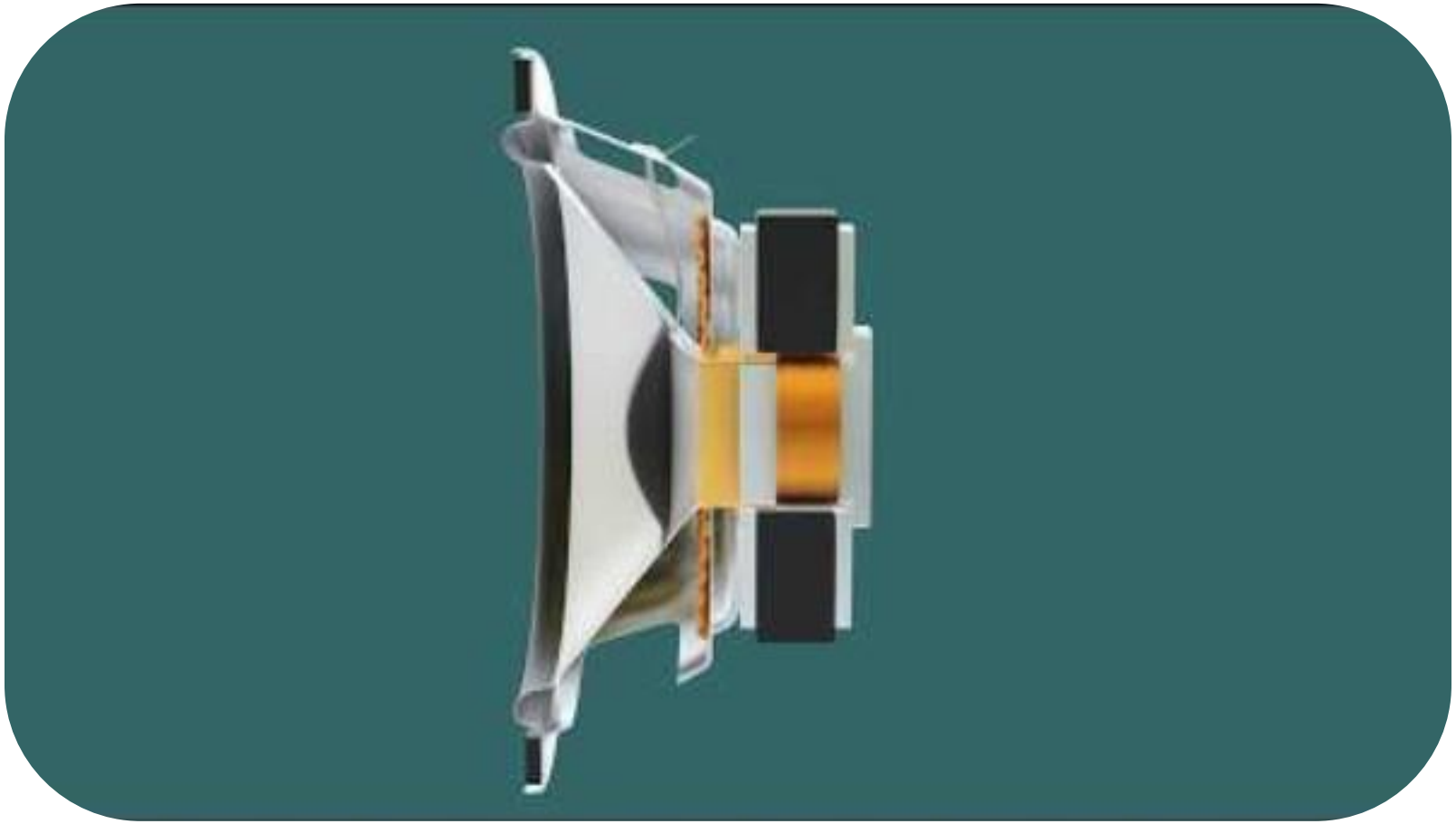
- **Bit Rate**: **Amount of data transferred per second** (unit: bits/sec)
  - Example: **320K MP3** files → **320kbps** (320,000 bits per second)
  - Example: **YouTube** recommendation → **128 kbps** for mono and **384 kbps** for stereo
  - Determines the file size!

📖 Reading: Microphones: Measuring Sound Pressure

youtu.be/d_crXXbuEKE

Into The Ordinary, "How do microphones work? - Recording Explained," *YouTube*, June 24, 2015.

📖 Reading: Speakers: Reproducing Sound Pressure



youtu.be/RxdFP31QYAg

Animagraffs, "How Speakers Make Sound," *YouTube*, December 16, 2020.

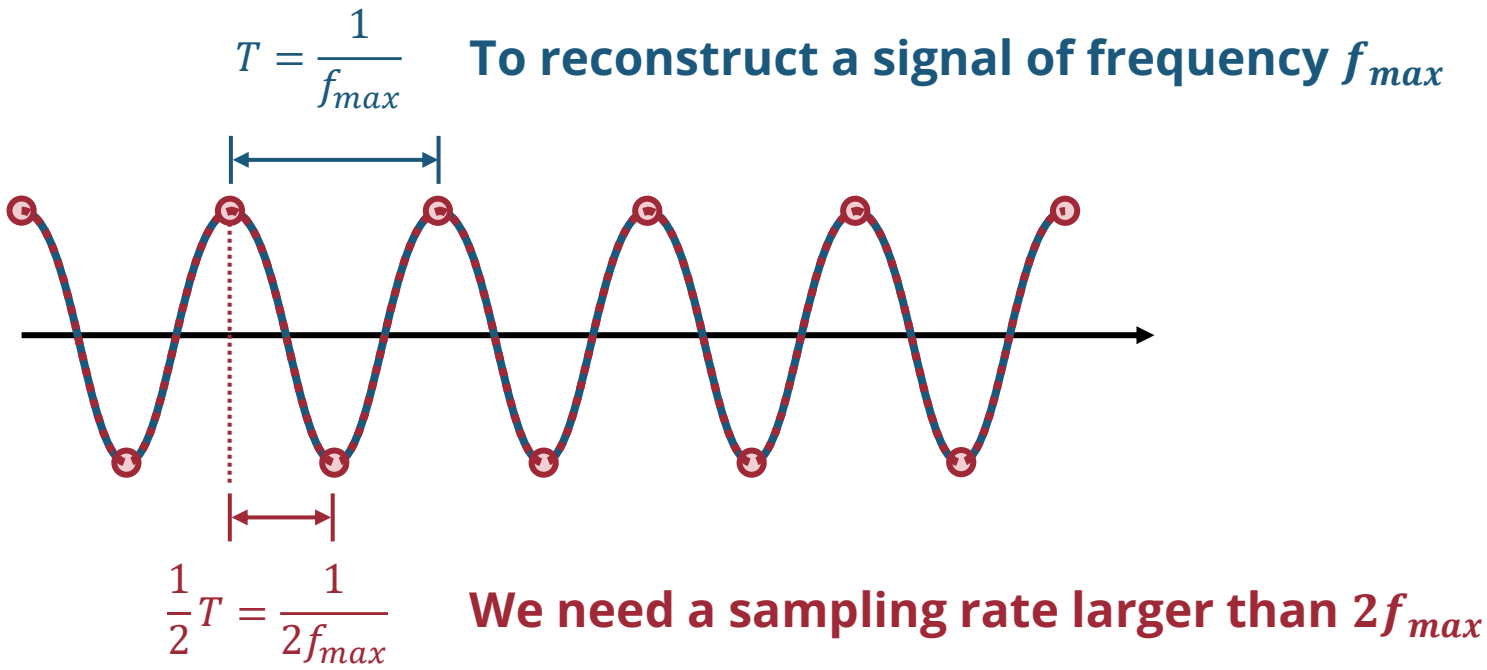
📖 Reading: Evolution of Music Media

youtu.be/-bVketPj5to

Sampling Theorem

Nyquist–Shannon Sampling Theorem

- **Theorem**: If a signal contains no frequencies higher than f_{max}, then the signal can be perfectly reconstructed when sampled at a rate $f_s > 2f_{max}$
  - $2f_{max}$ is usually referred to as the **Nyquist rate**

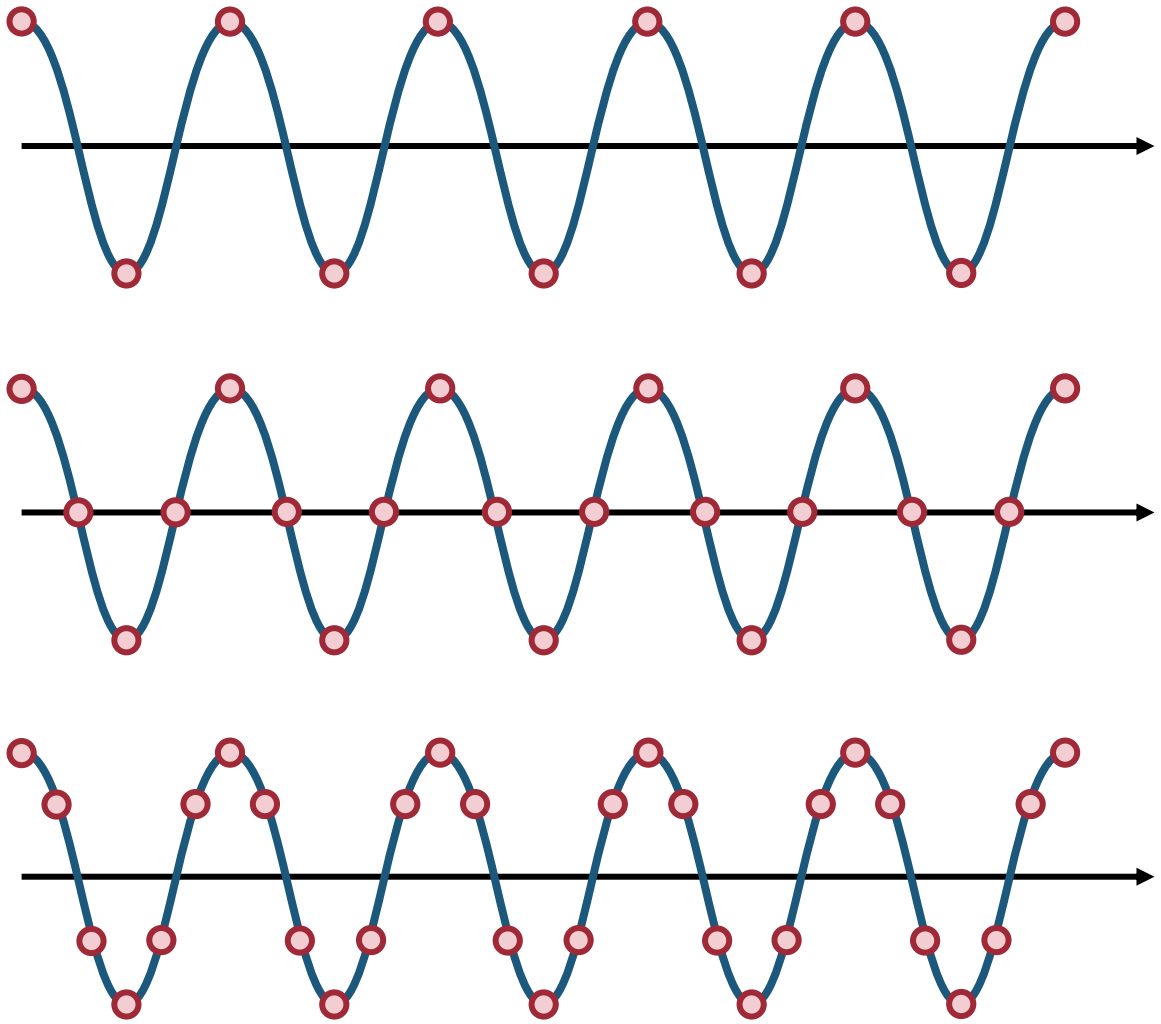


Sampling Theorem: Oversampling

Critically sampled
$(f_s = 2f_{max})$

Oversampled
$(f_s = 4f_{max})$

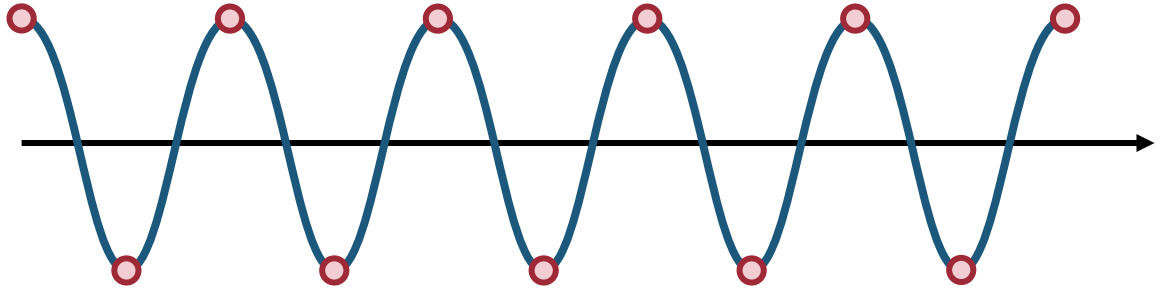
Oversampled
$(f_s = 6f_{max})$



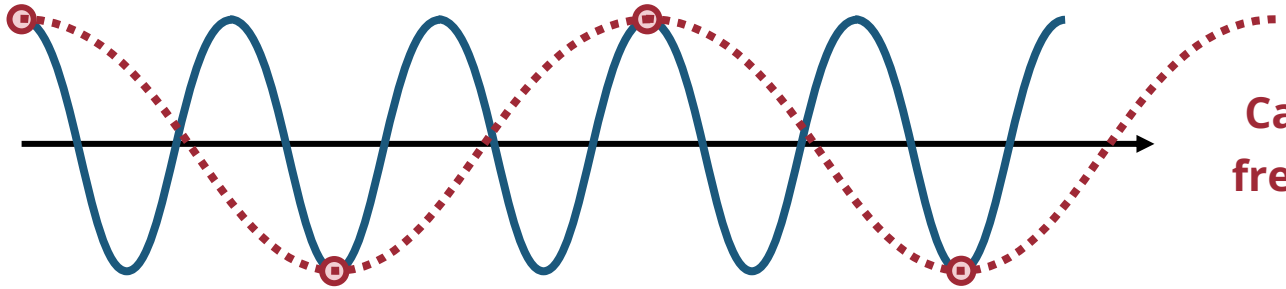
Reconstruction is possible!

Sampling Theorem: Undersampling

Critically sampled
$(f_s = 2f_{max})$

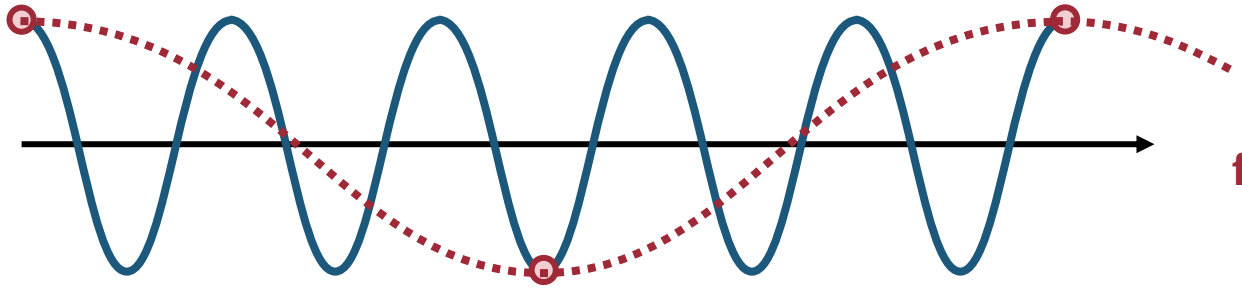


Undersampled
$(f_s = \frac{2}{3} f_{max})$



Can only reconstruct frequency up to $\frac{1}{3} f_{max}$

Undersampled
$(f_s = \frac{2}{5} f_{max})$



Can only reconstruct frequency up to $\frac{1}{3} f_{max}$

🤔 Sampling Theorem

- **Telephone audio** is sampled at **8 kHz**. What is the maximum frequency it can reconstruct?
  - **4 kHz**

- To cover the **human hearing range of 20 Hz to 20 kHz**, what is the minimum sampling rate required?
  - **40 kHz**

Sampling Rate & Frequency Range

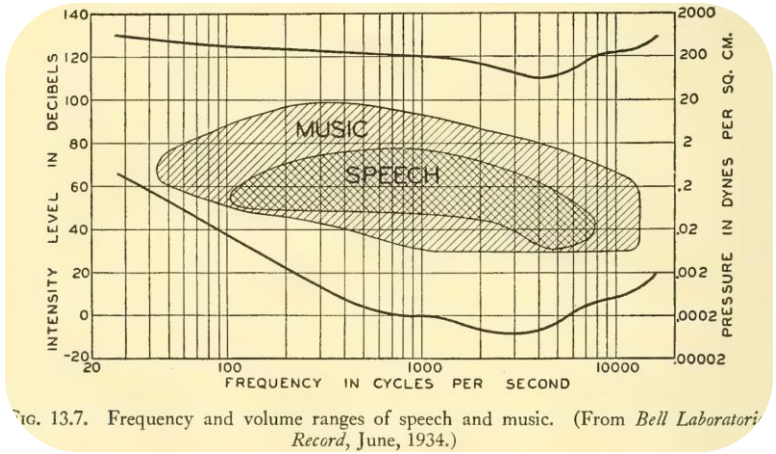
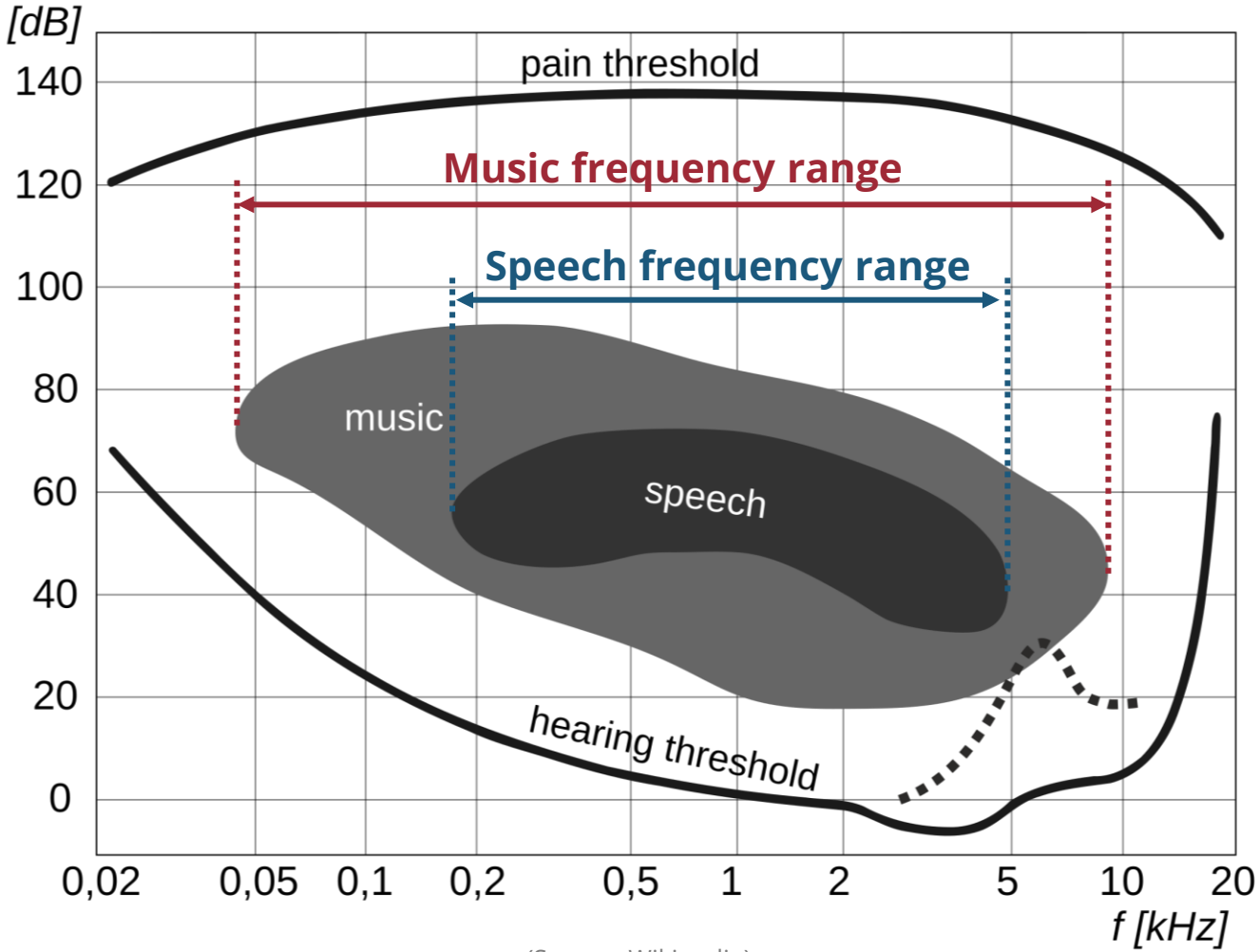


Fig. 13.7. Frequency and volume ranges of speech and music. (From *Bell Laboratori Record*, June, 1934.)

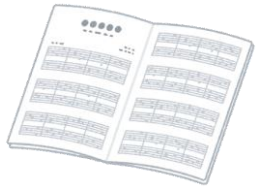
(Source: Bell Laboratories Record 1934 & Olson 1947)



(Source: Wikipedia)

Bell Laboratories Record, 12(6):314, 1934.
Harry Ferdinand Olson, "Speech, Music and Hearing," *Elements of acoustical engineering Hardcover*, p. 326, 1947.
en.wikipedia.org/wiki/Hearing_range

Recap

Music Representations

Symbolic music representations

Text-based

```
Program_change_0,
Note_on_60, Time_shift_2, Note_off_60,
Note_on_60, Time_shift_2, Note_off_60,
Note_on_76, Time_shift_2, Note_off_67,
Note_on_67, Time_shift_2, Note_off_67,
...
```

MIDI

Image-based

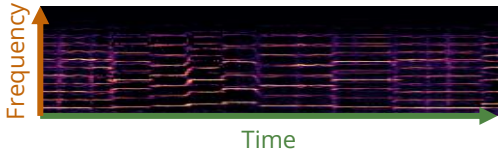
Piano roll

Audio-domain music representations

Time series-based

Waveform

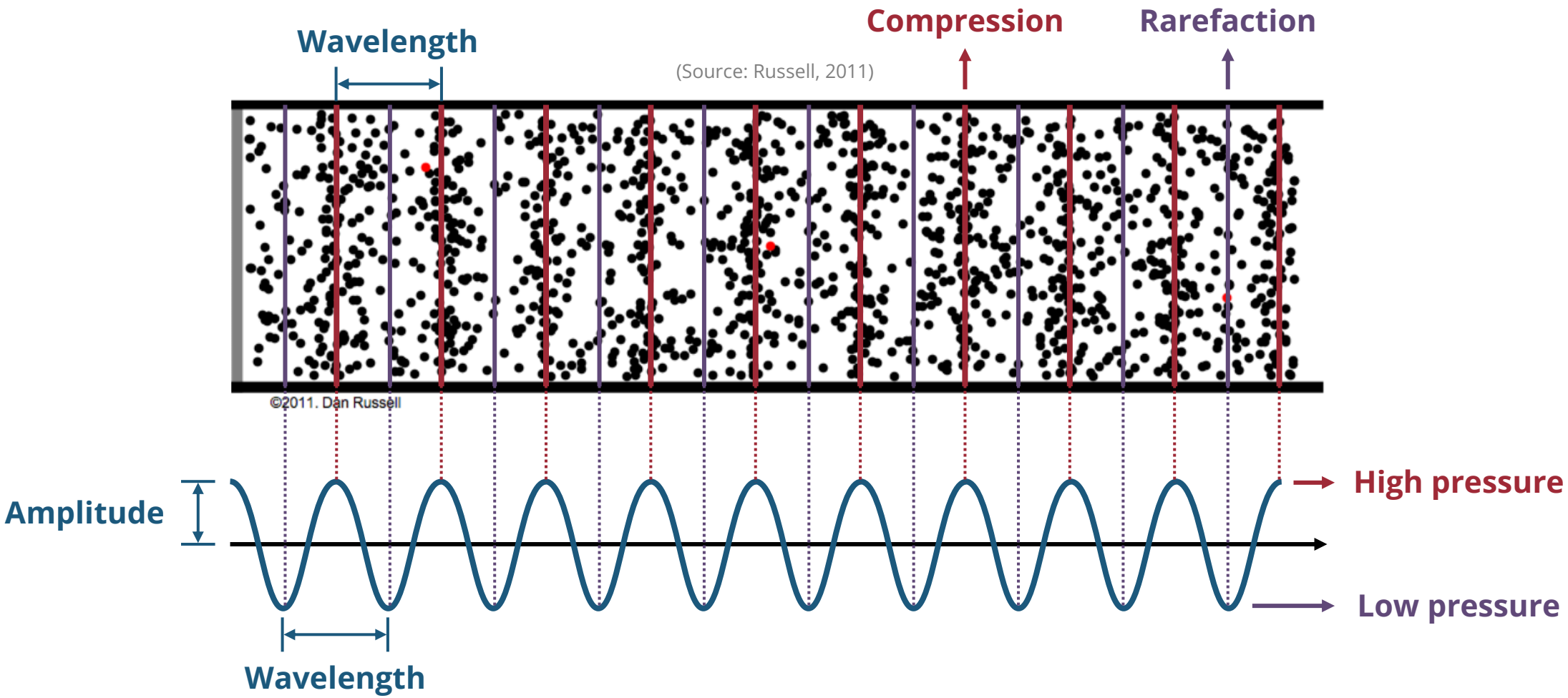
Image-based



Spectrogram

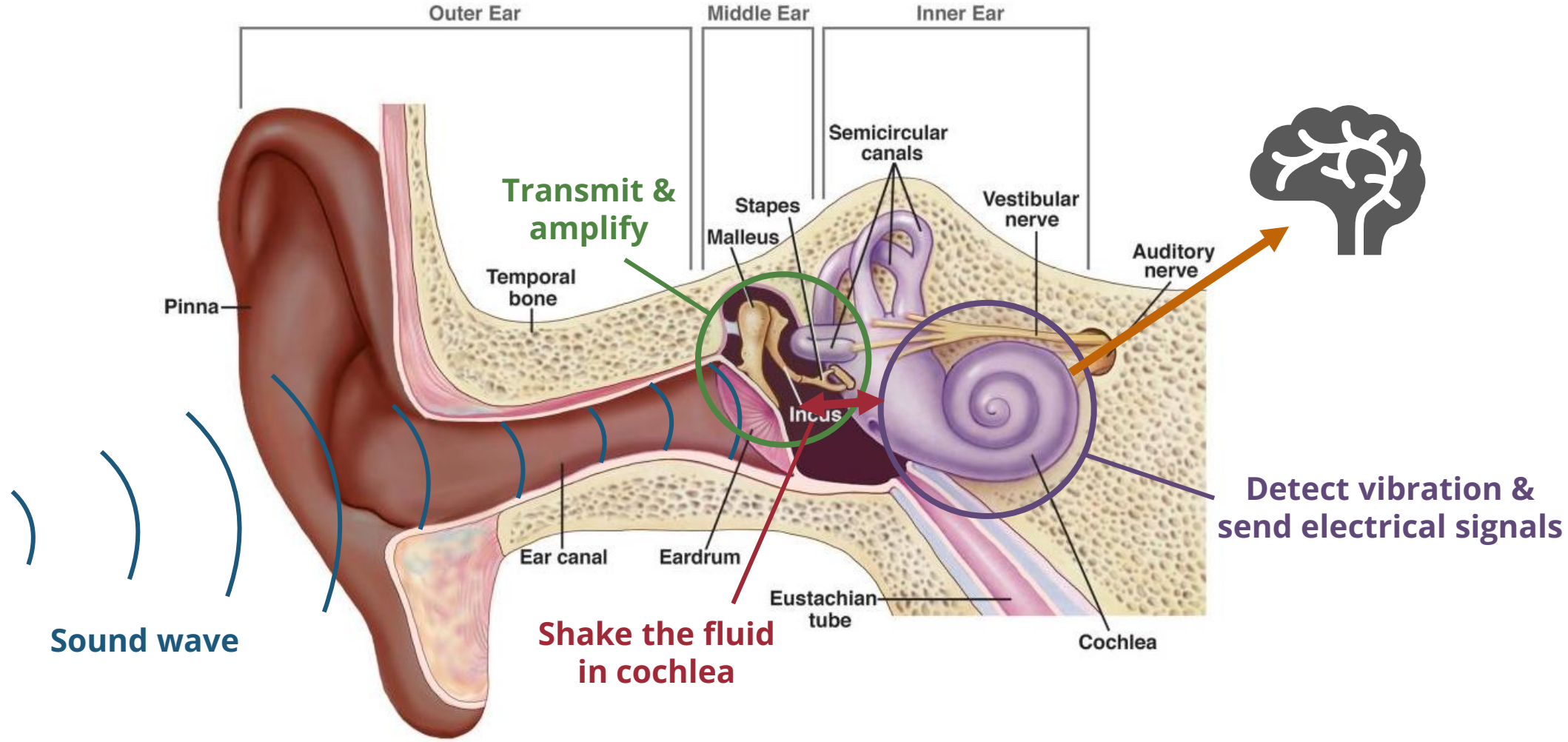
Today's topic!

Longitudinal vs. Transverse Waves



Dan Russell, "Longitudinal and Transverse Wave Motion," *Acoustics and Vibration Animations*, 1998 (updated on 2016).

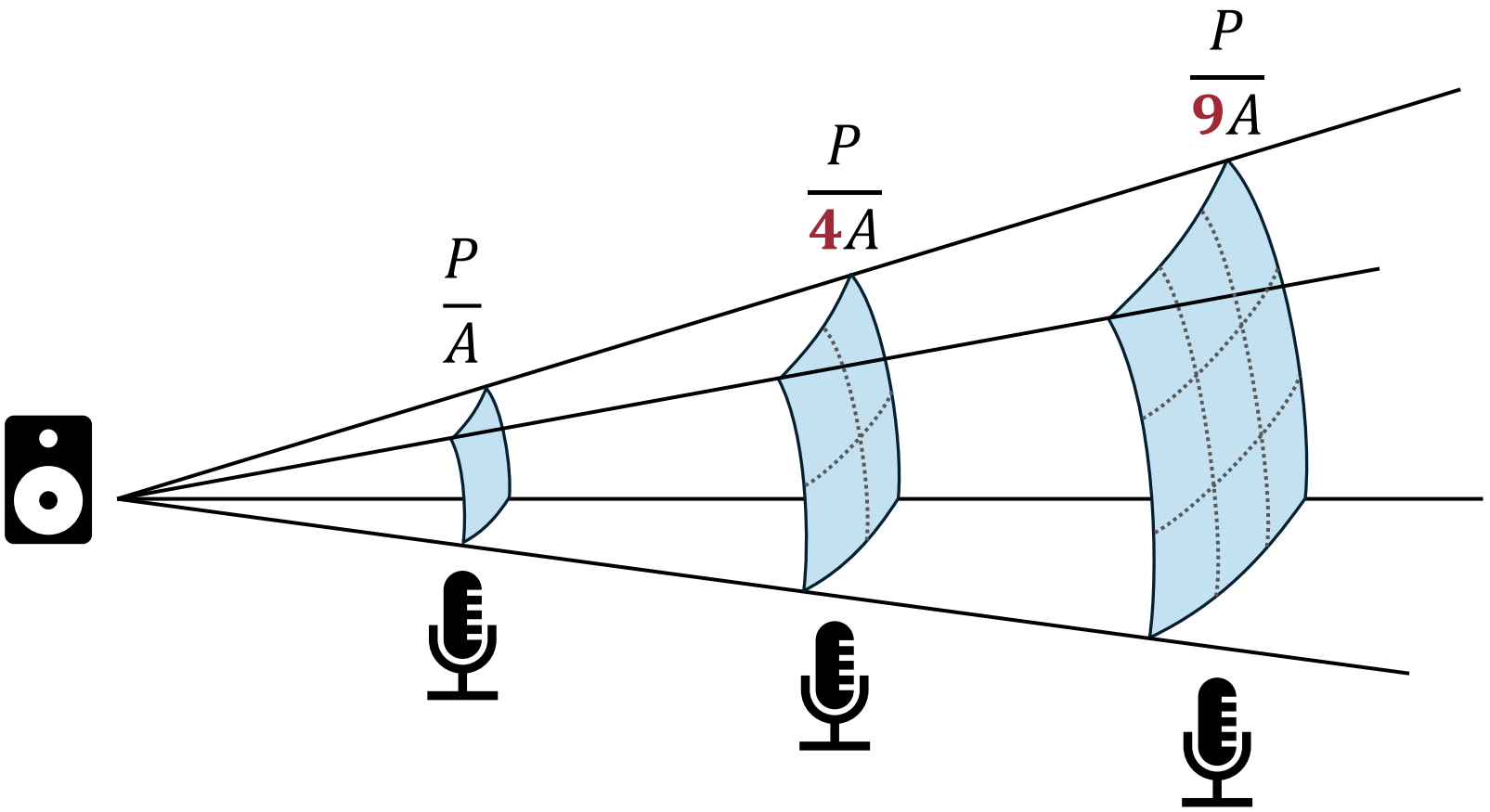
Human Ears



(Source: NIH/NIDCD)

National Institute on Deafness and Other Communication Disorders (NIDCD), "How Do We Hear?," National Institutes of Health (NIH), 2015.

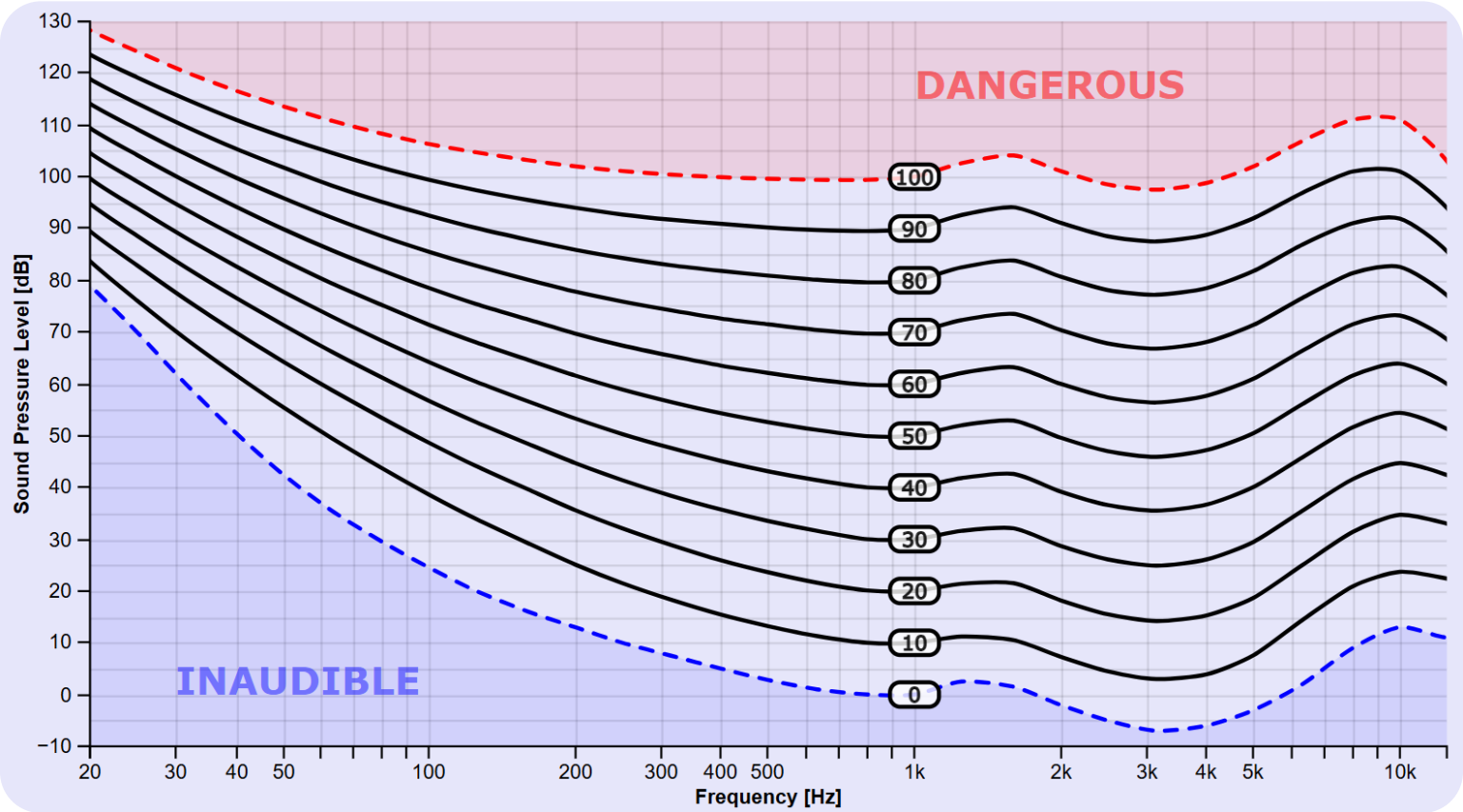
Sound Propagation & Inverse Square Law



Inverse Square Law

$$I \propto \frac{1}{r^2}$$

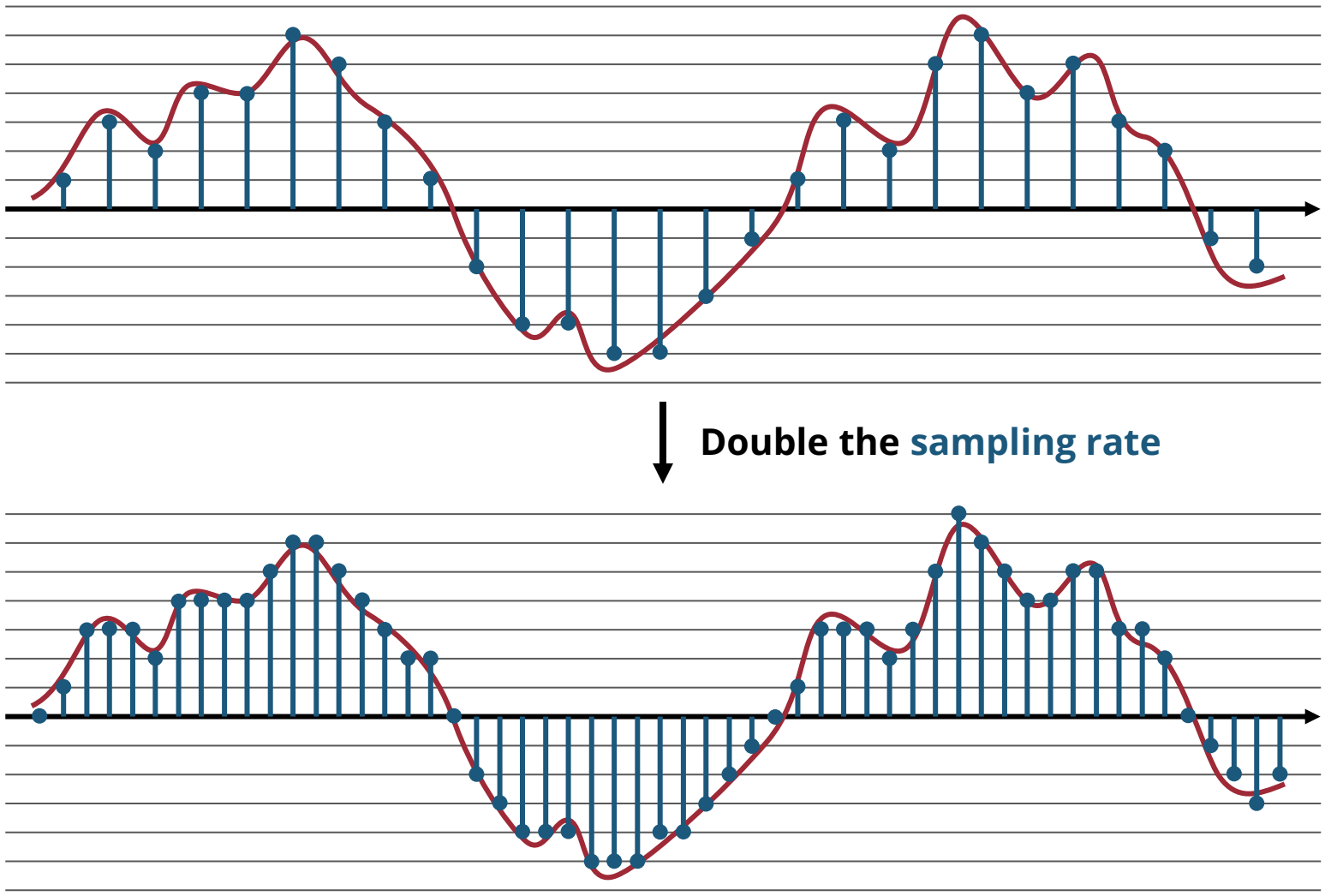
Loudness Perception: Equal-loudness Contours



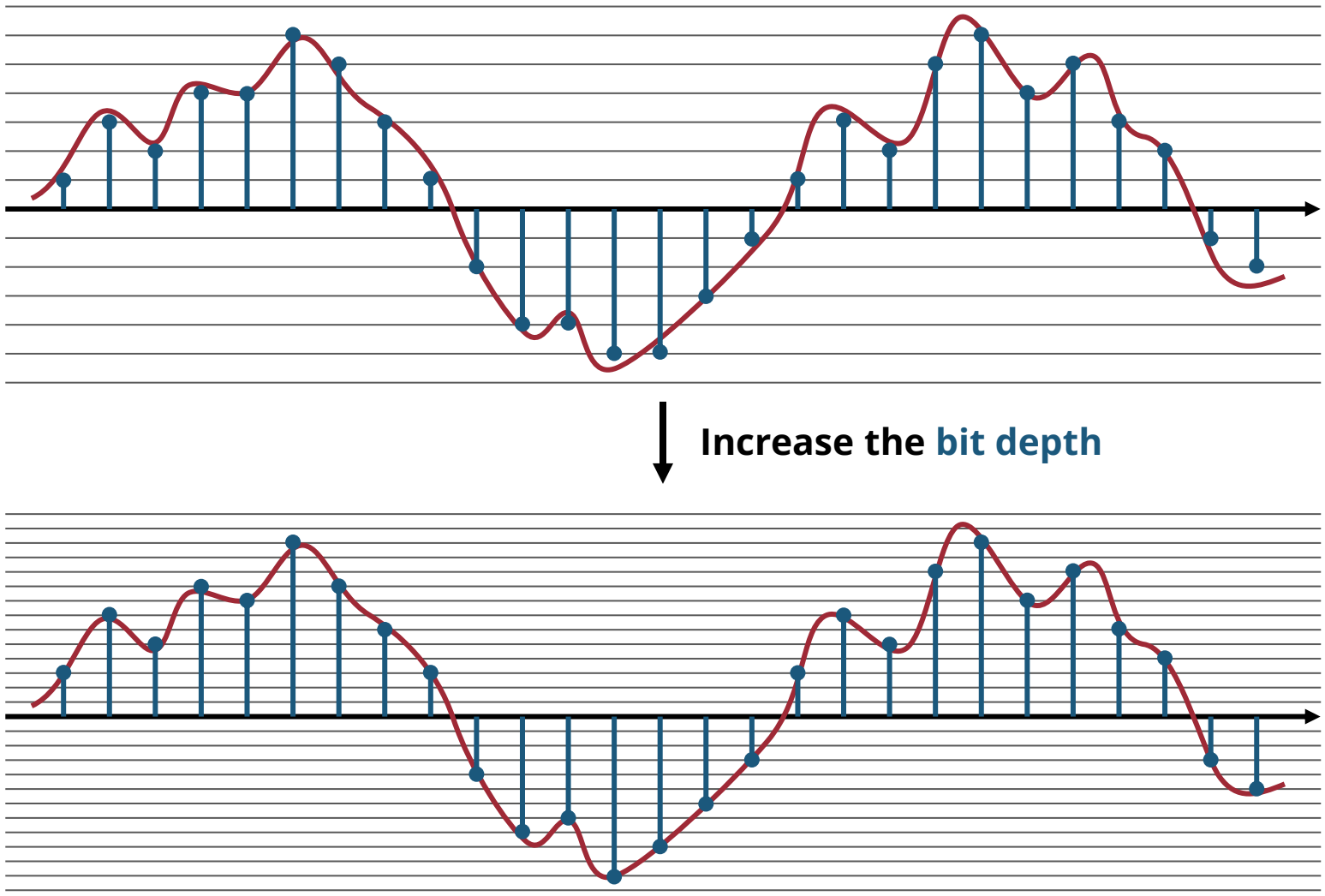
(Source: Parker, 2024)

Noah J. Parker, "Normal Equal-Loudness-Level Contours," *Math of Sound*, 2024.

Sampling Rate: Temporal Resolution

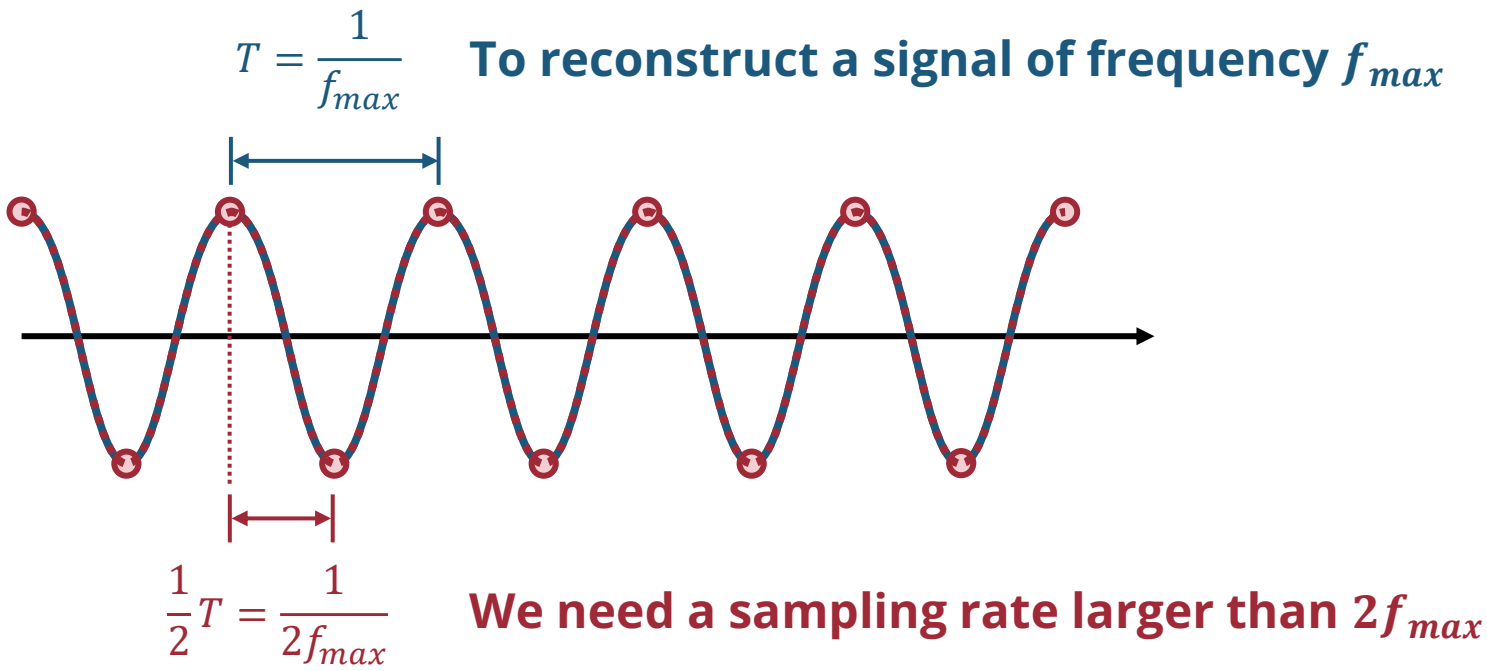


Bit Depth: Dynamic Resolution



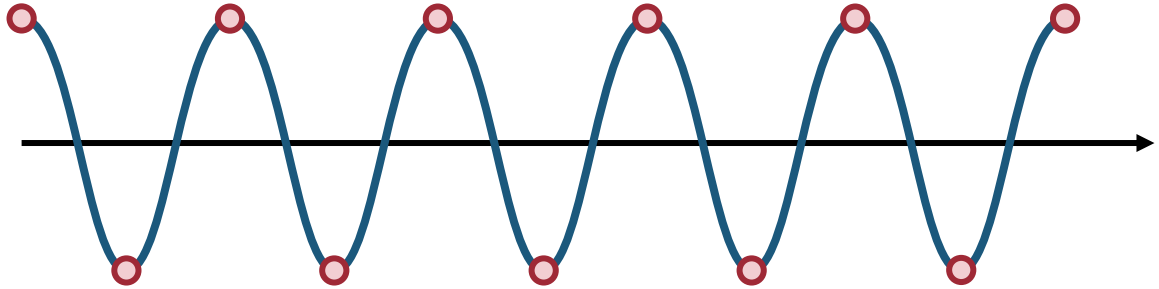
Nyquist–Shannon Sampling Theorem

- **Theorem**: If a signal contains no frequencies higher than f_{max}, then the signal can be perfectly reconstructed when sampled at a rate $f_s > 2f_{max}$
  - $2f_{max}$ is usually referred to as the **Nyquist rate**

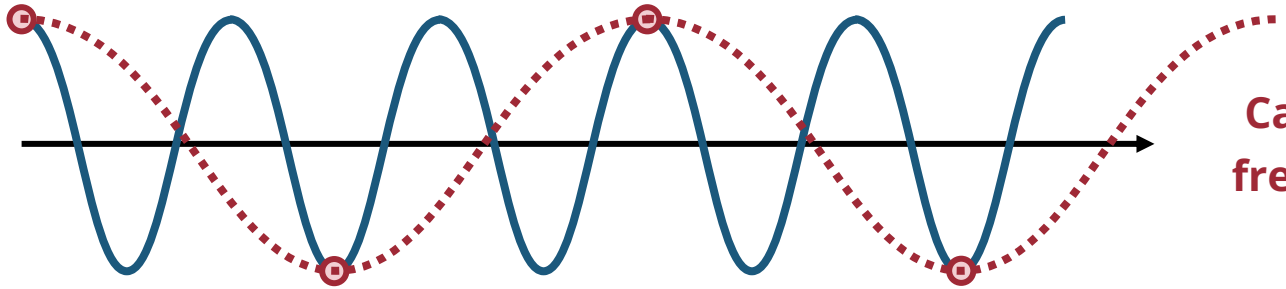


Sampling Theorem: Undersampling

Critically sampled
$(f_s = 2f_{max})$

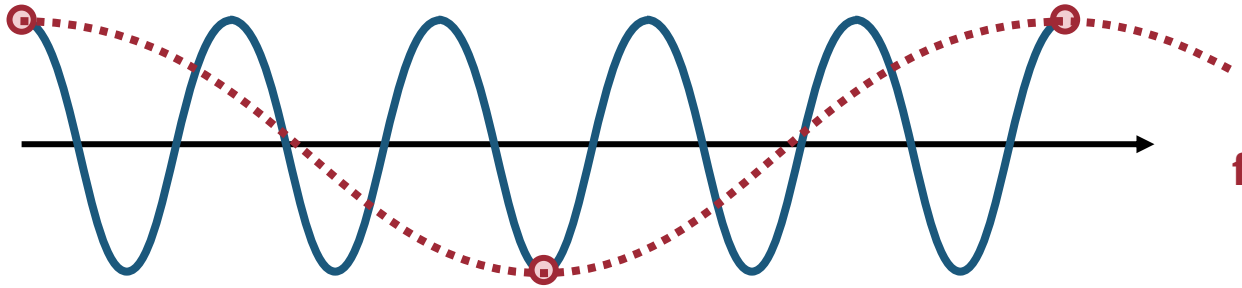


Undersampled
$(f_s = \frac{2}{3} f_{max})$



Can only reconstruct frequency up to $\frac{1}{3} f_{max}$

Undersampled
$(f_s = \frac{2}{5} f_{max})$

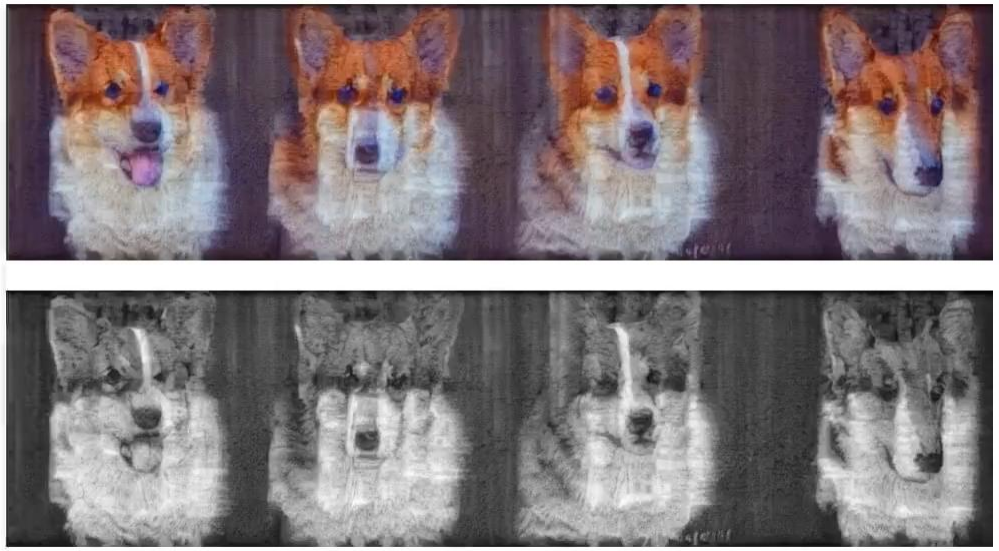


Can only reconstruct frequency up to $\frac{1}{3} f_{max}$

Next Lecture

Spectral Analysis

Image prompt: a colorful photo of corgis



Audio prompt: dog barking

(Source: Chen et al., 2024)