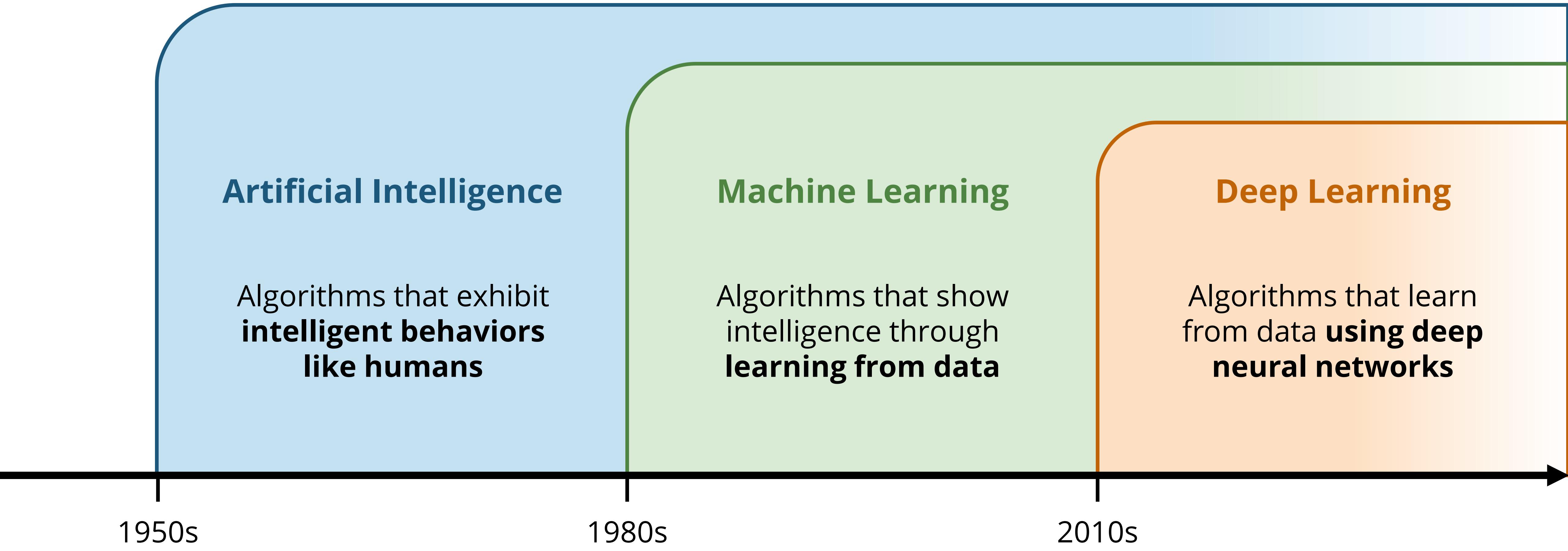
PAT 498/598 (Winter 2025)

Music & AI

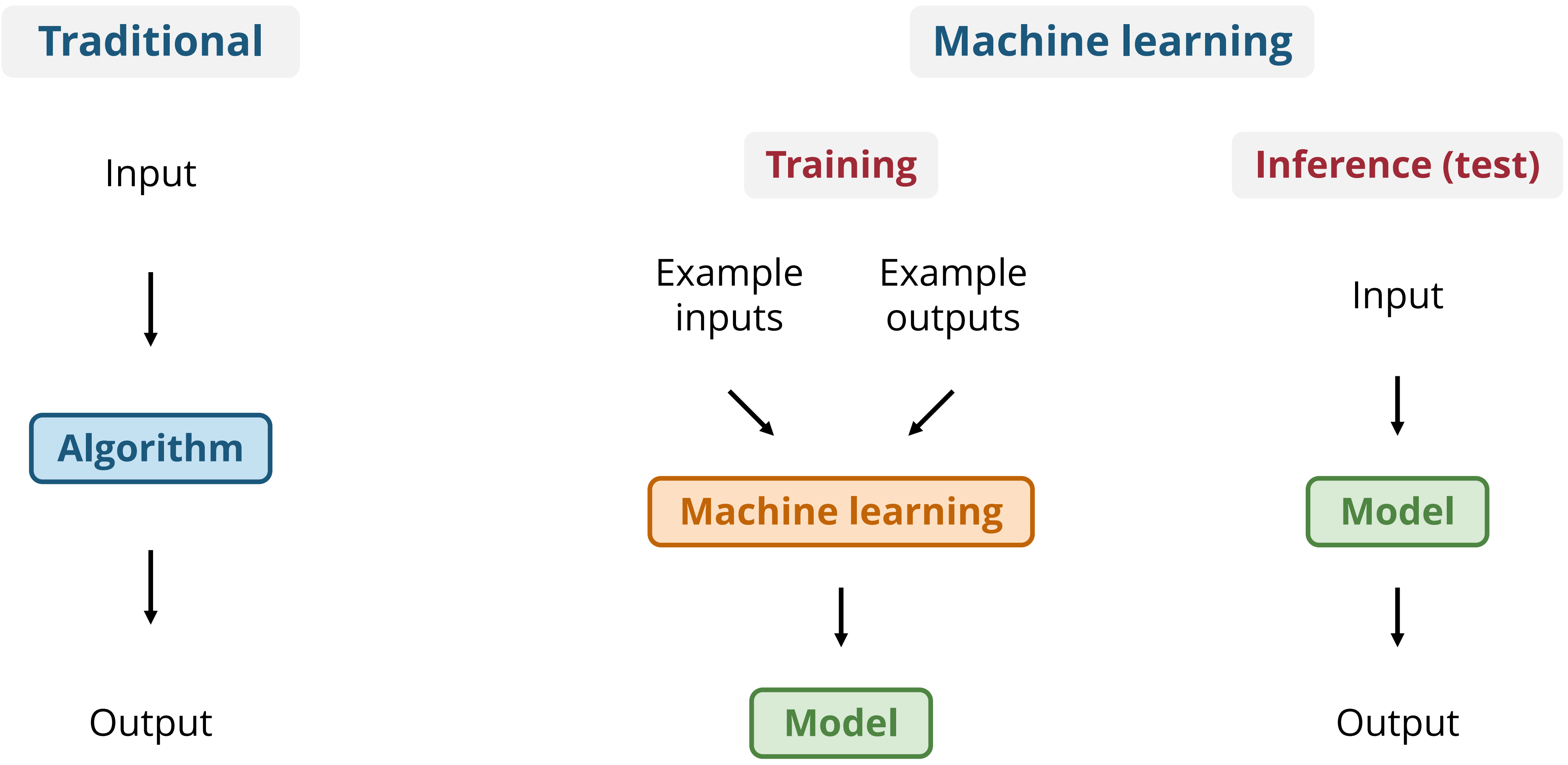
Lecture 5: Music & Audio Processing Fundamentals

Instructor: Hao-Wen Dong

(Recap) AI vs ML vs DL



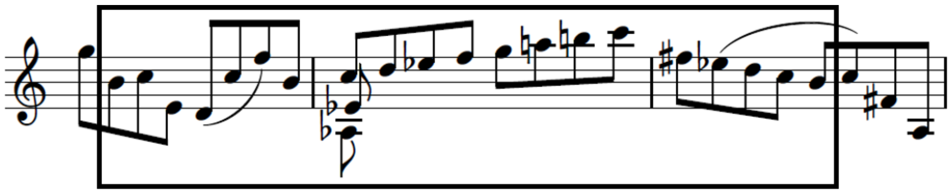
(Recap) Machine Learning



(Recap) Components of a Machine Learning Model

Improve on task T,
with respect to performance metric P,
based on experience E

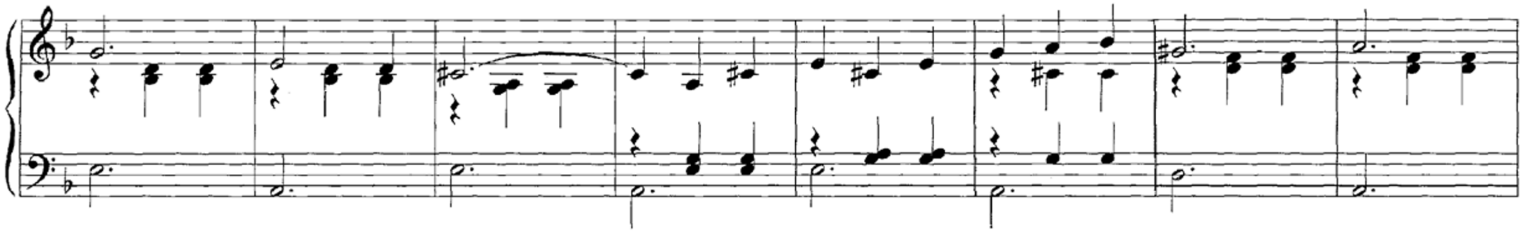
- **Task T** — **Violin transcription**
- **Performance metric P** — Percentage of correctly predicted notes
- **Experience E** — Recordings with sheet music



Hao-Wen Dong, Cong Zhou, Taylor Berg-Kirkpatrick, and Julian McAuley, "Deep Performer: Score-to-Audio Music Performance Synthesis," ICASSP, 2022.

(Recap) Components of a Machine Learning Model

Improve on task T,
with respect to performance metric P,
based on experience E

- **Task T** — **Beat tracking**
- **Performance metric P** — Average difference from actual timings
- **Experience E** — Recordings with beat timestamps

(Source: Müller)

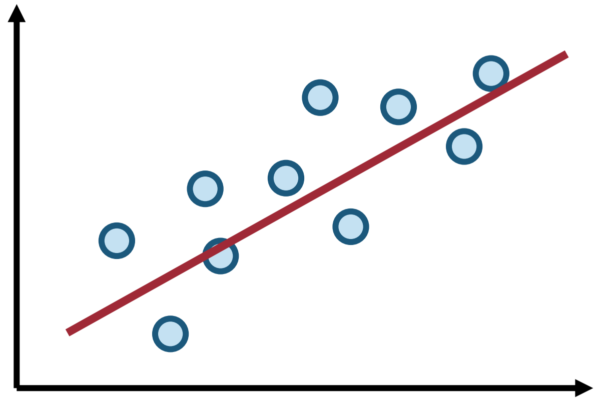
audiolabs-erlangen.de/resources/MIR/FMP/C6/C6S3_BeatTracking.html
Meinard Müller, "Fundamentals of Music Processing – Using Python and Jupyter Notebooks," *Springer Verlag*, 2021.
Meinard Müller and Frank Zalkow, "FMP Notebooks: Educational Material for Teaching and Learning Fundamentals of Music Processing," *ISMIR*, 2019.

(Recap) Types of Machine Learning

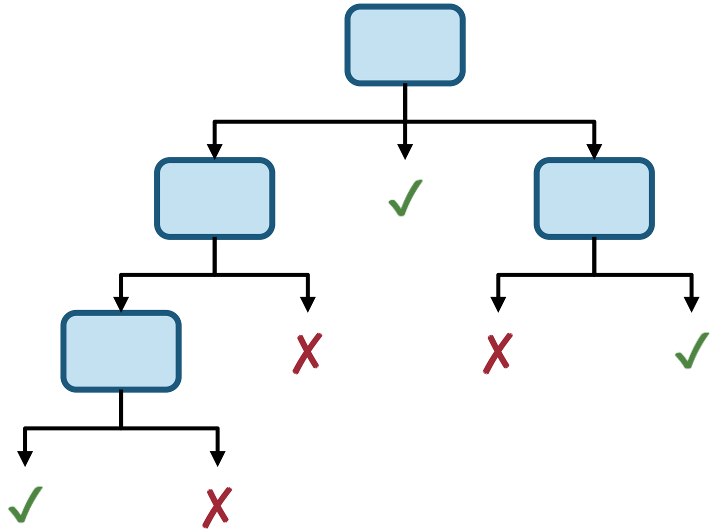
- **Supervised learning** — Given **pairs of example inputs and outputs**
  - **Classification**: *discrete* outputs
  - **Regression**: *continuous* outputs
- **Unsupervised learning** — Given ***only* example inputs**
  - **Self-supervised learning**
- **Semi-supervised learning** — Given **example inputs** and **a few example outputs**
- **Reinforcement learning** — Given **scalar rewards** for **a sequence of actions**

(Recap) Examples of Machine Learning Algorithms

Linear regression



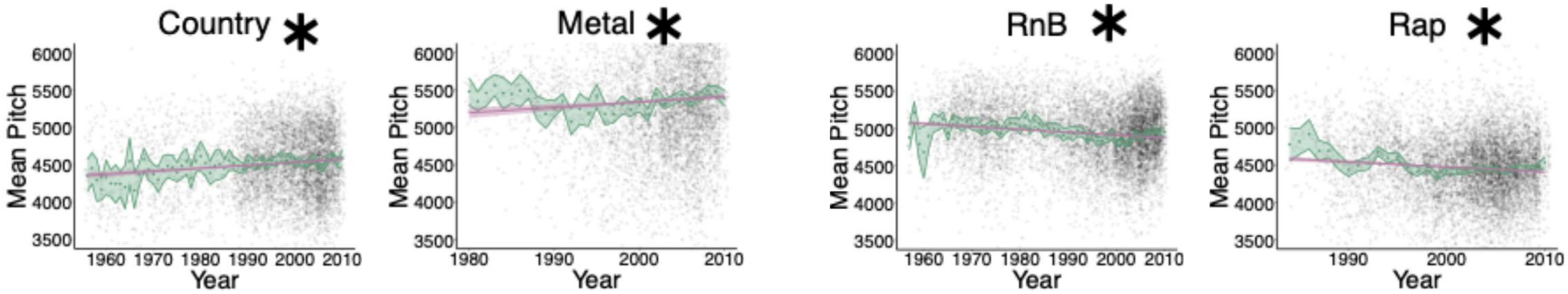
Decision tree



(Recap) Example: Linear Regression

Positive correlation

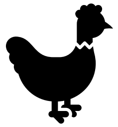
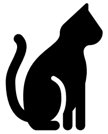
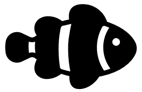
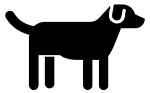
Negative correlation

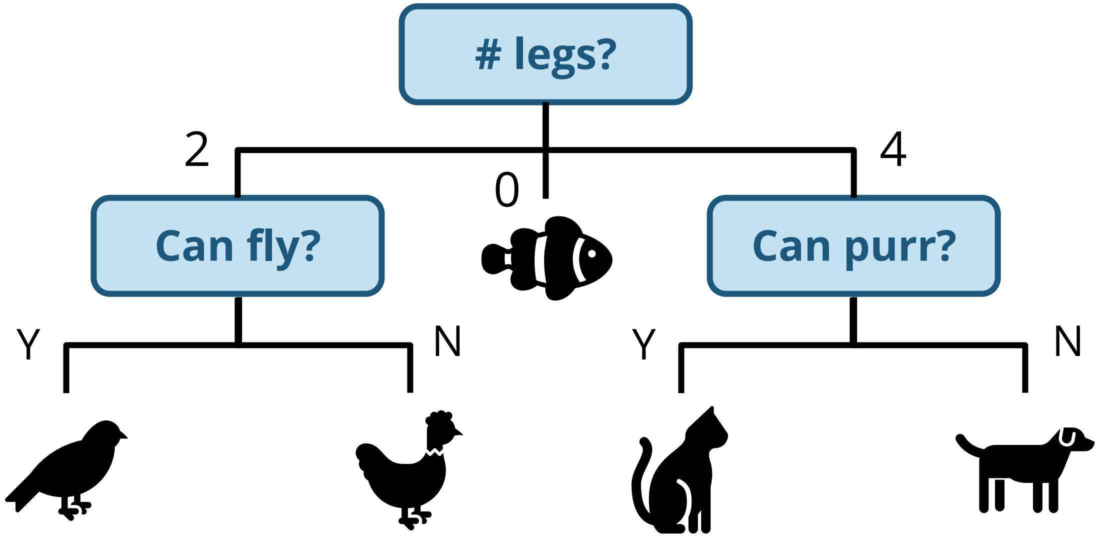


(Source: Georgieva et al., 2024)

Elena Georgieva, Pablo Ripollés, and Brian McFee, "The Changing Sound of Music: An Exploratory Corpus Study of Vocal Trends Over Time," *ISMIR*, 2024.

(Recap) Building a Decision Tree

	Can fly?	Can swim?	# of legs	Can purr?
	N	N	2	N
	N	N	4	Y
	N	Y	0	N
	Y	N	2	N
	N	N	4	N



(Recap) Components of a Machine Learning Model

Optimization

Defining inputs & outputs

Improve on task T,

with respect to performance metric P,

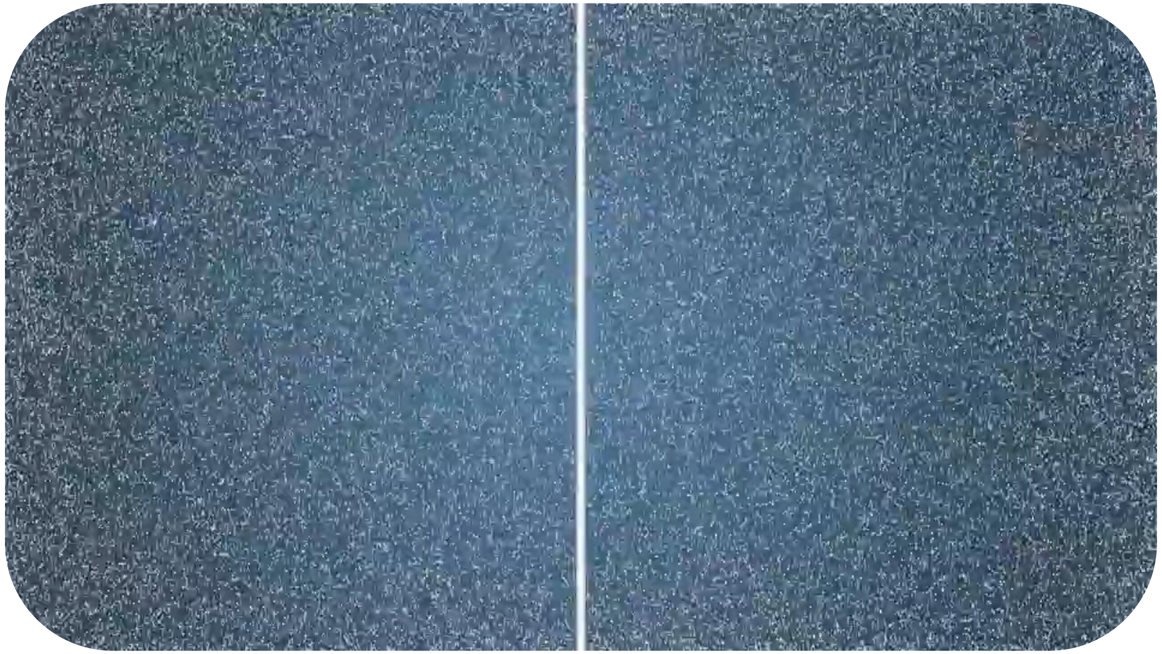
based on experience E

Objective function
(loss function)

Training data

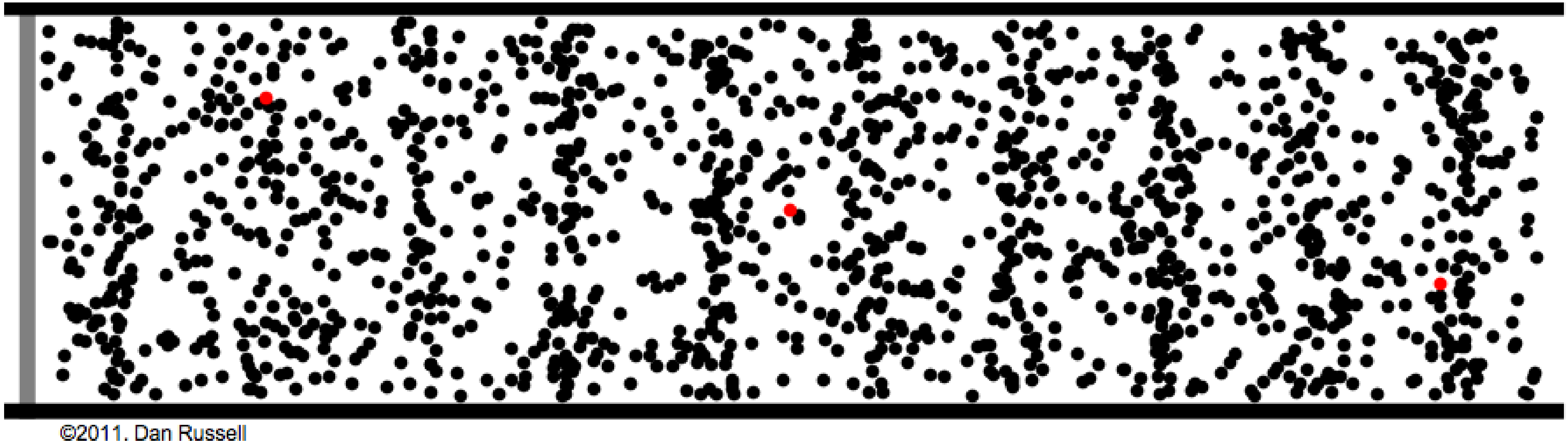
Music & Audio Processing

What is a Sound?



youtu.be/aPswnDcteS4

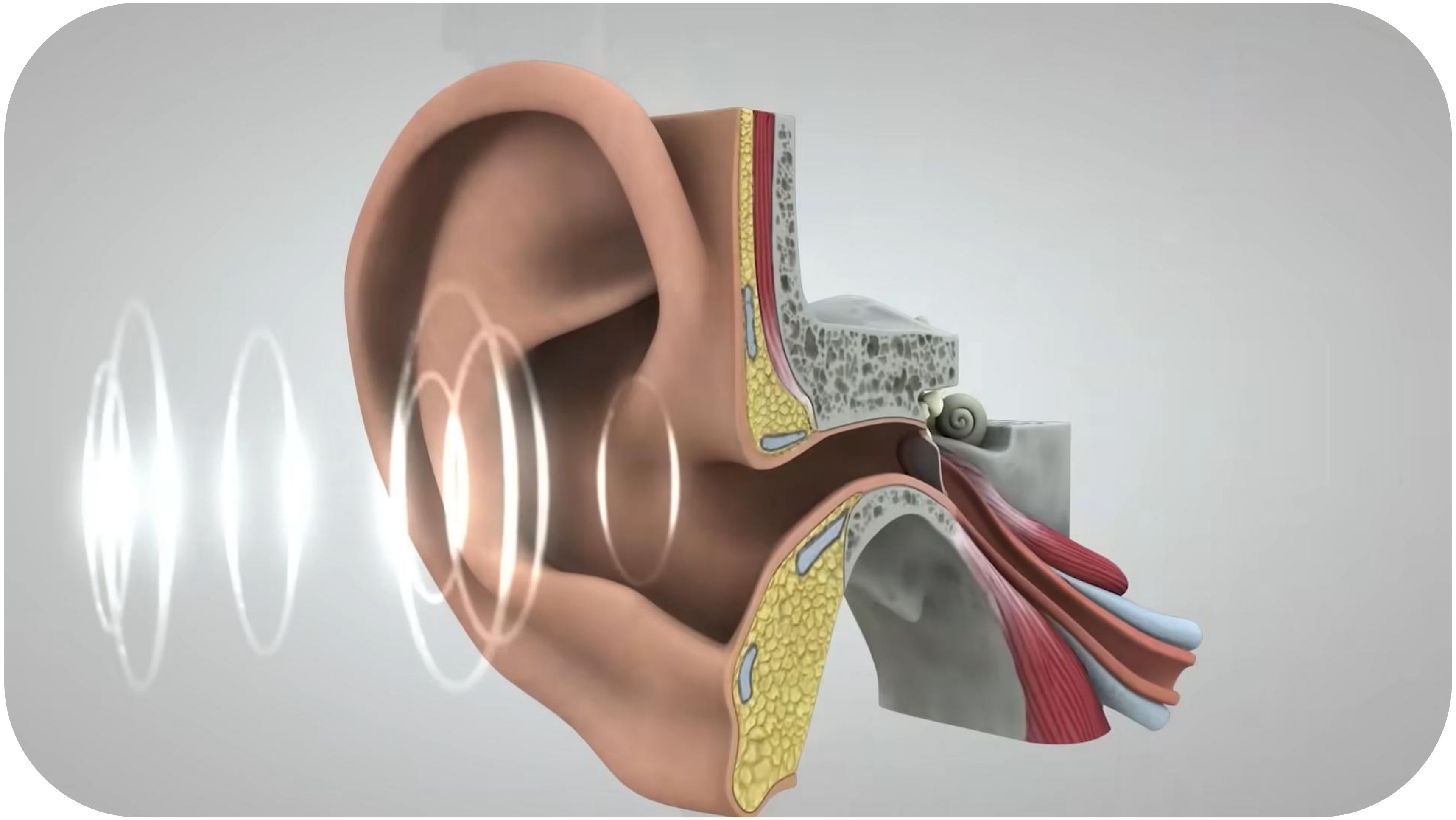
Sound is a Pressure Wave



(Source: Russell, 2011)

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

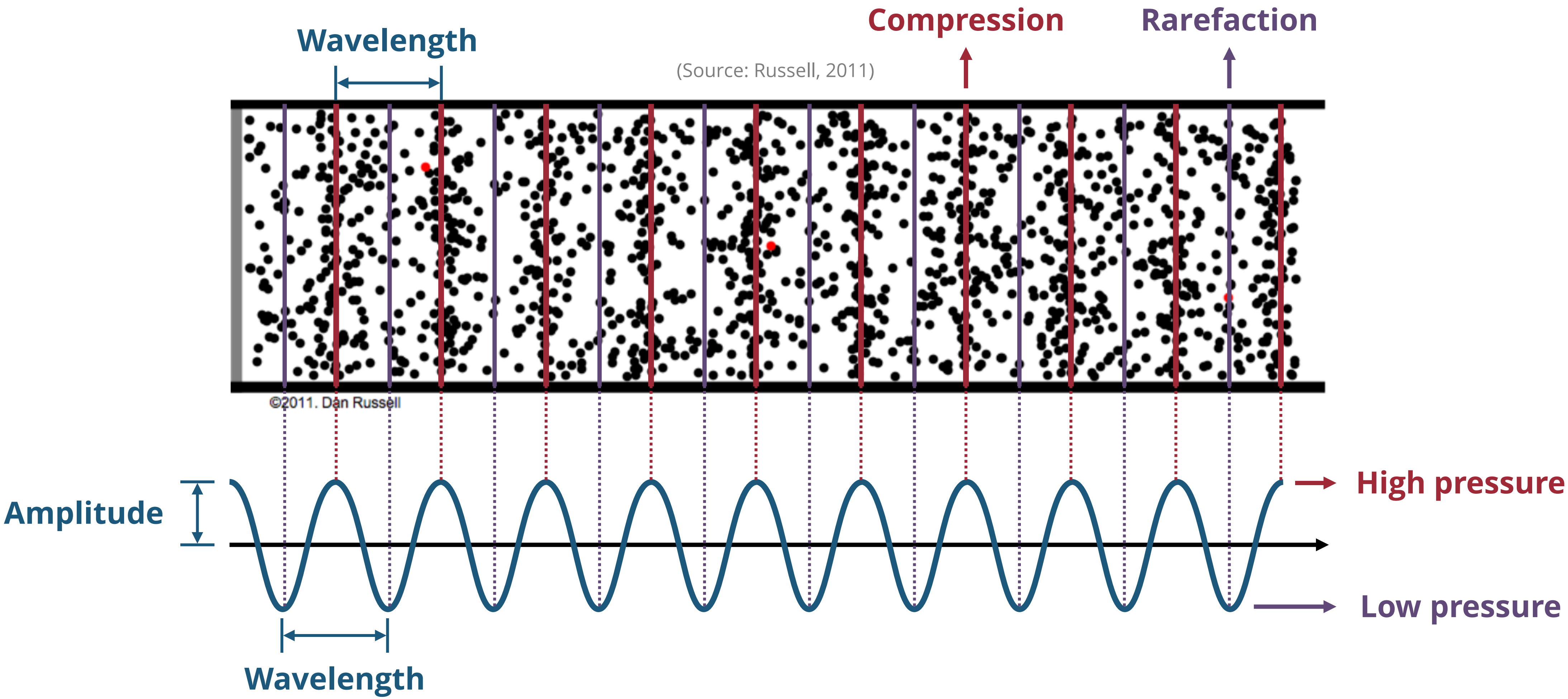
Human Auditory System



youtu.be/eQEaiZ2j9oc

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

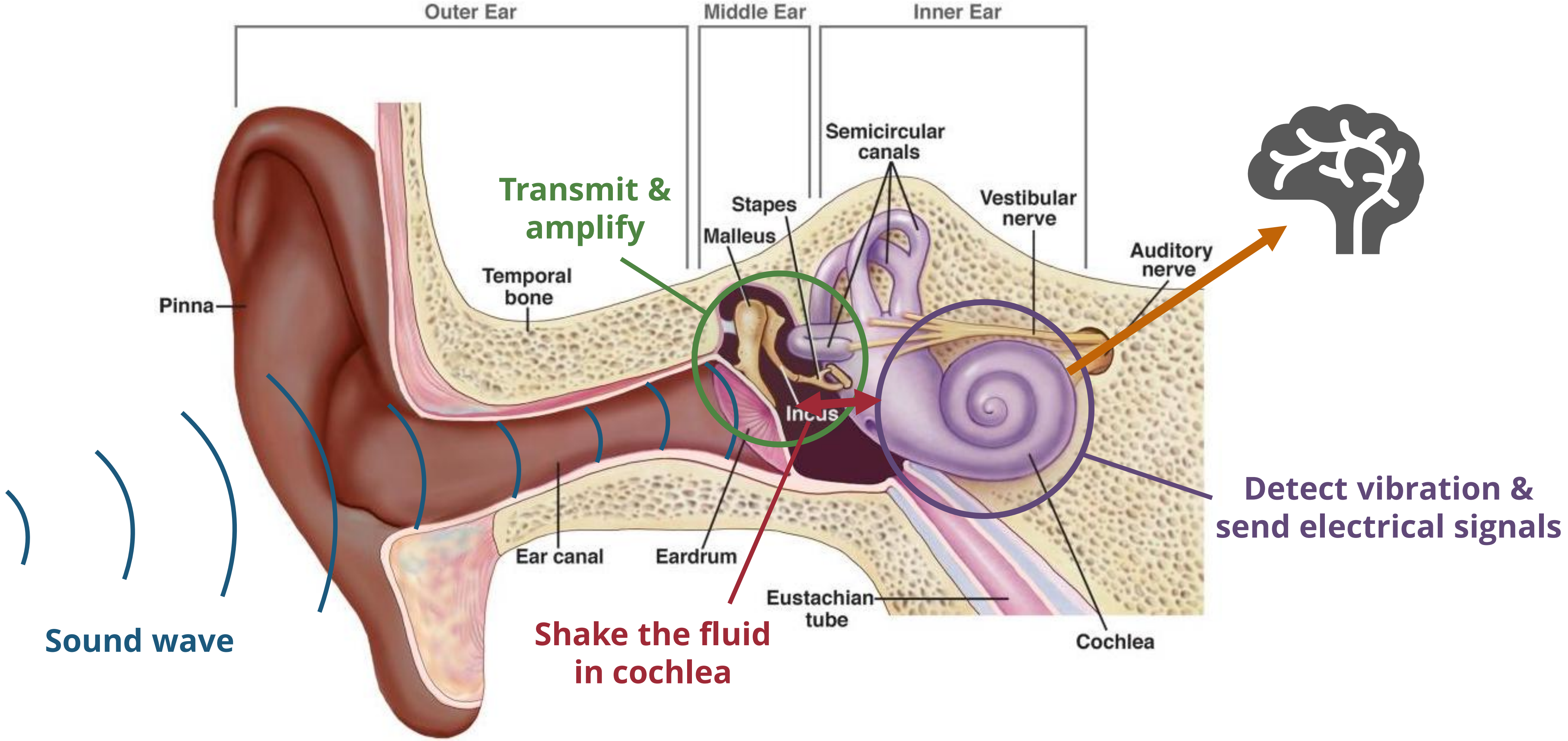
Longitudinal vs Transverse Waves



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

Auditory Perception

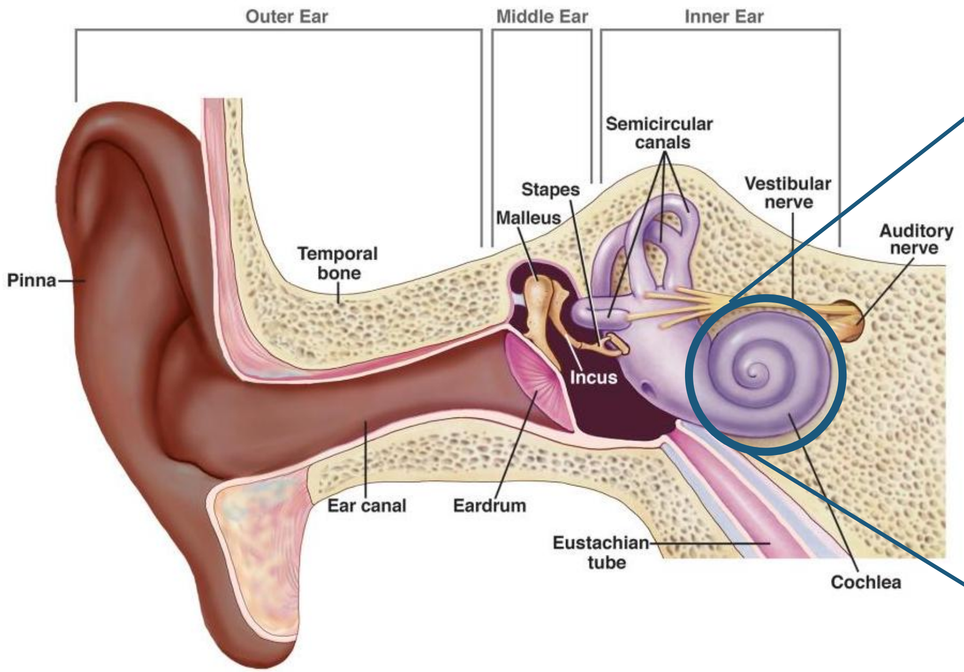
Human Ears



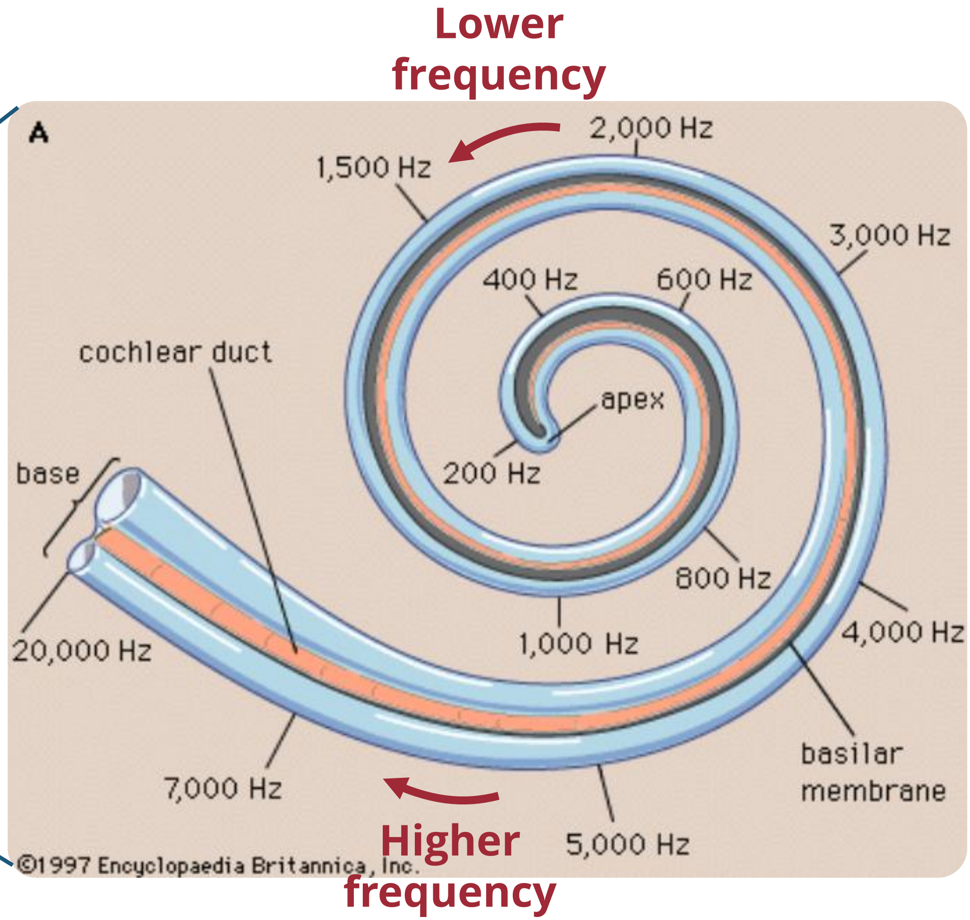
(Source: NIH/NIDCD)

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

Cochlea in the Inner Ear



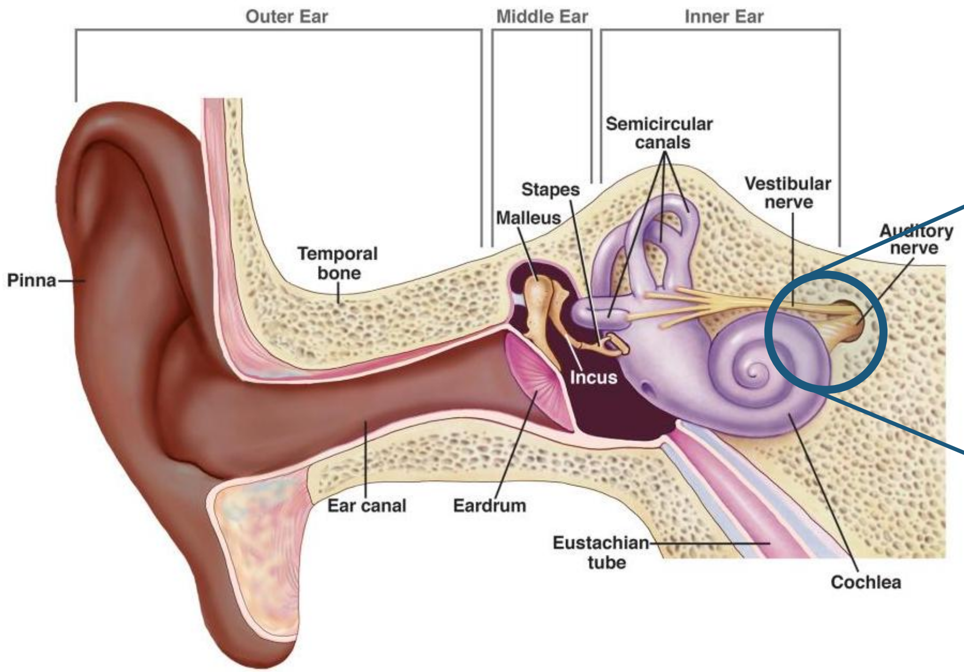
(Source: NIH/NIDCD)



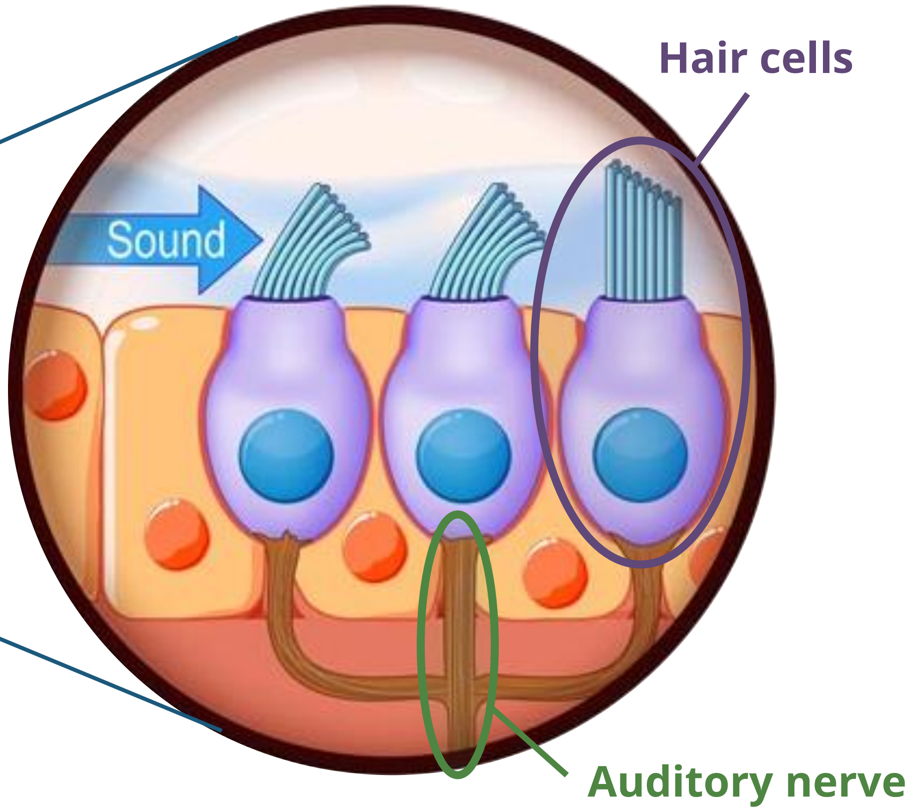
(Source: Britannica)

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," *Britannica*.

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 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)

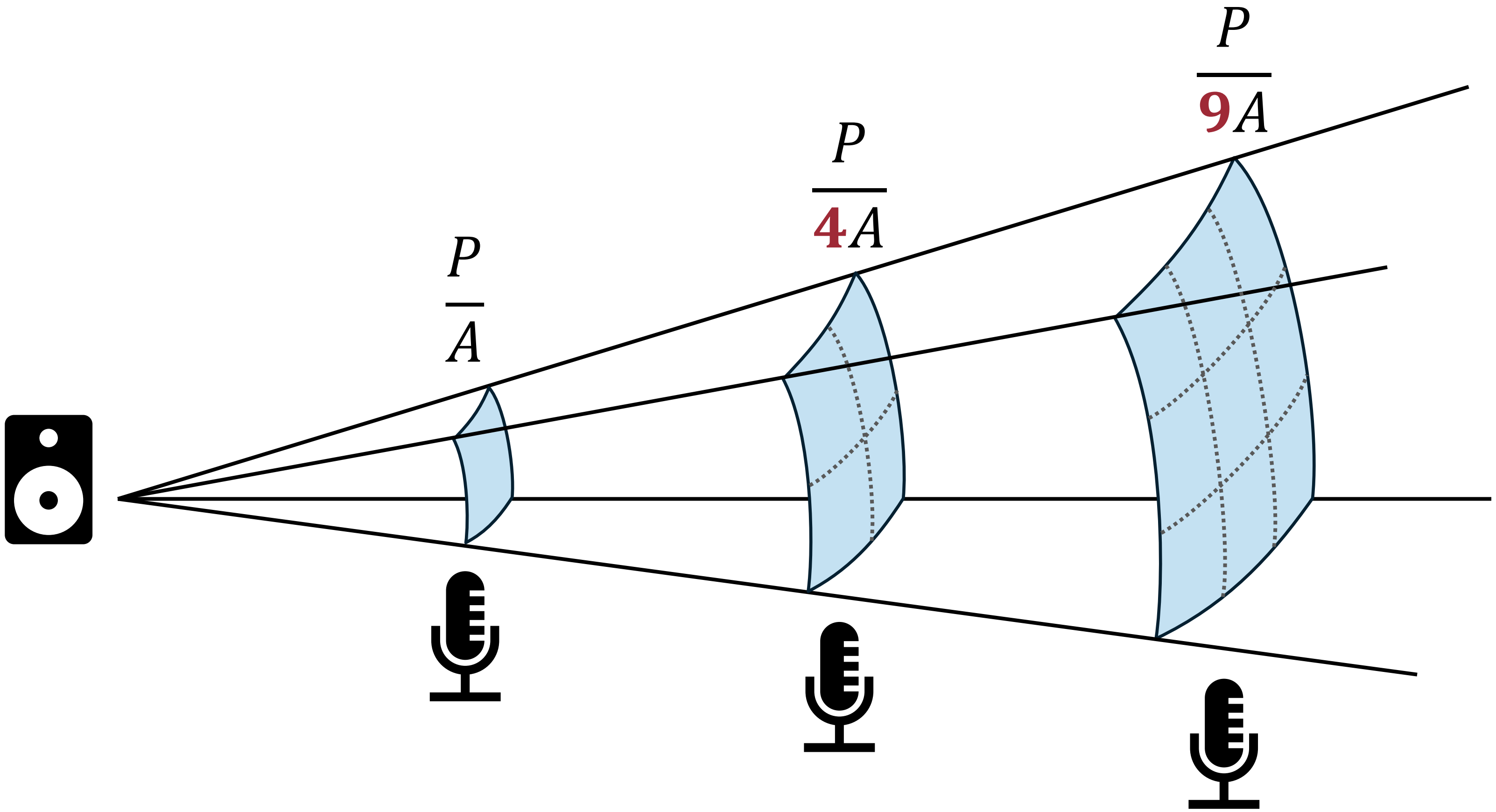
Richard E. Berg, "The decibel scale," *Britannica*.

Common Gains in Decibels

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

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

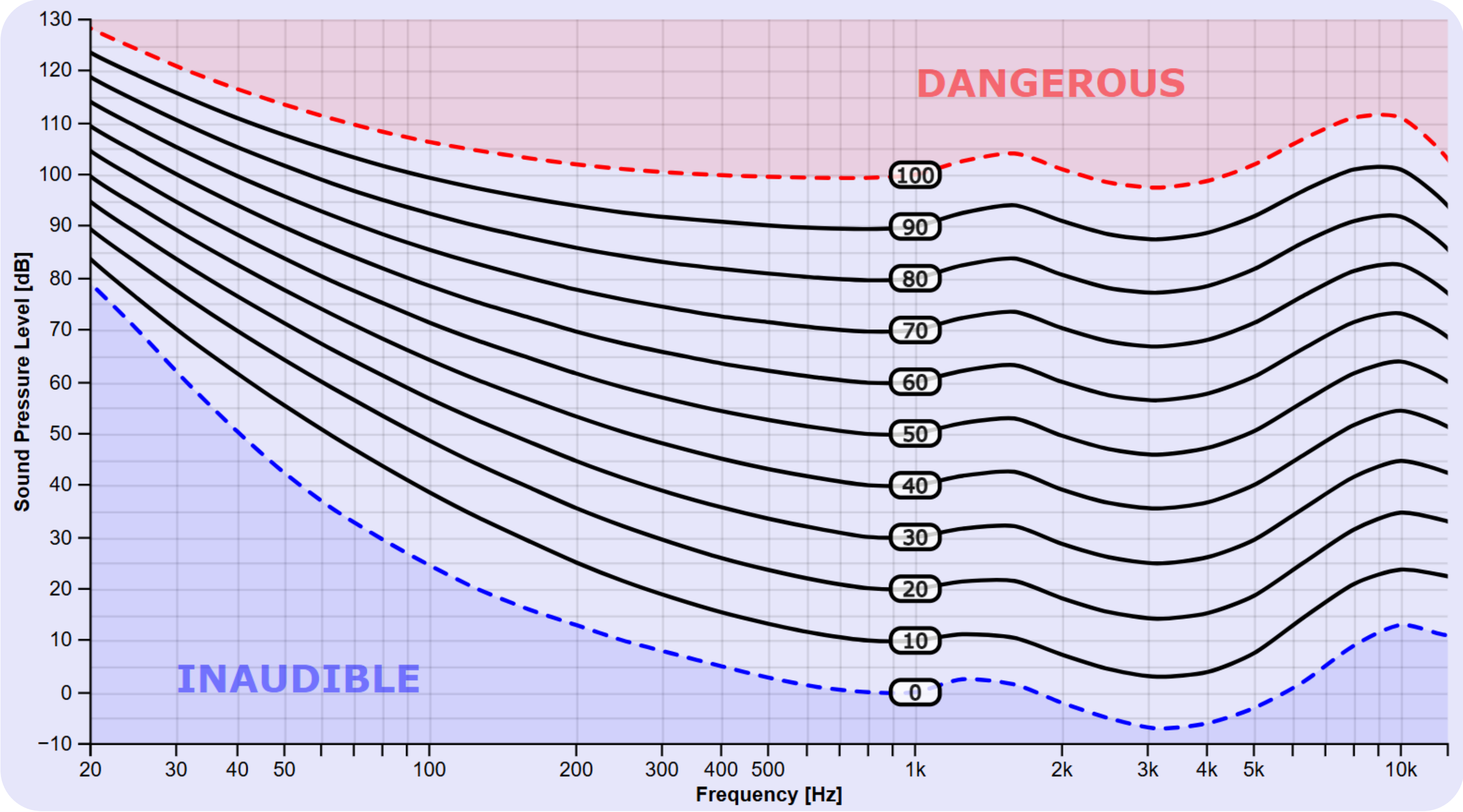
Sound Propagation & Inverse Square Law



Inverse Square Law

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

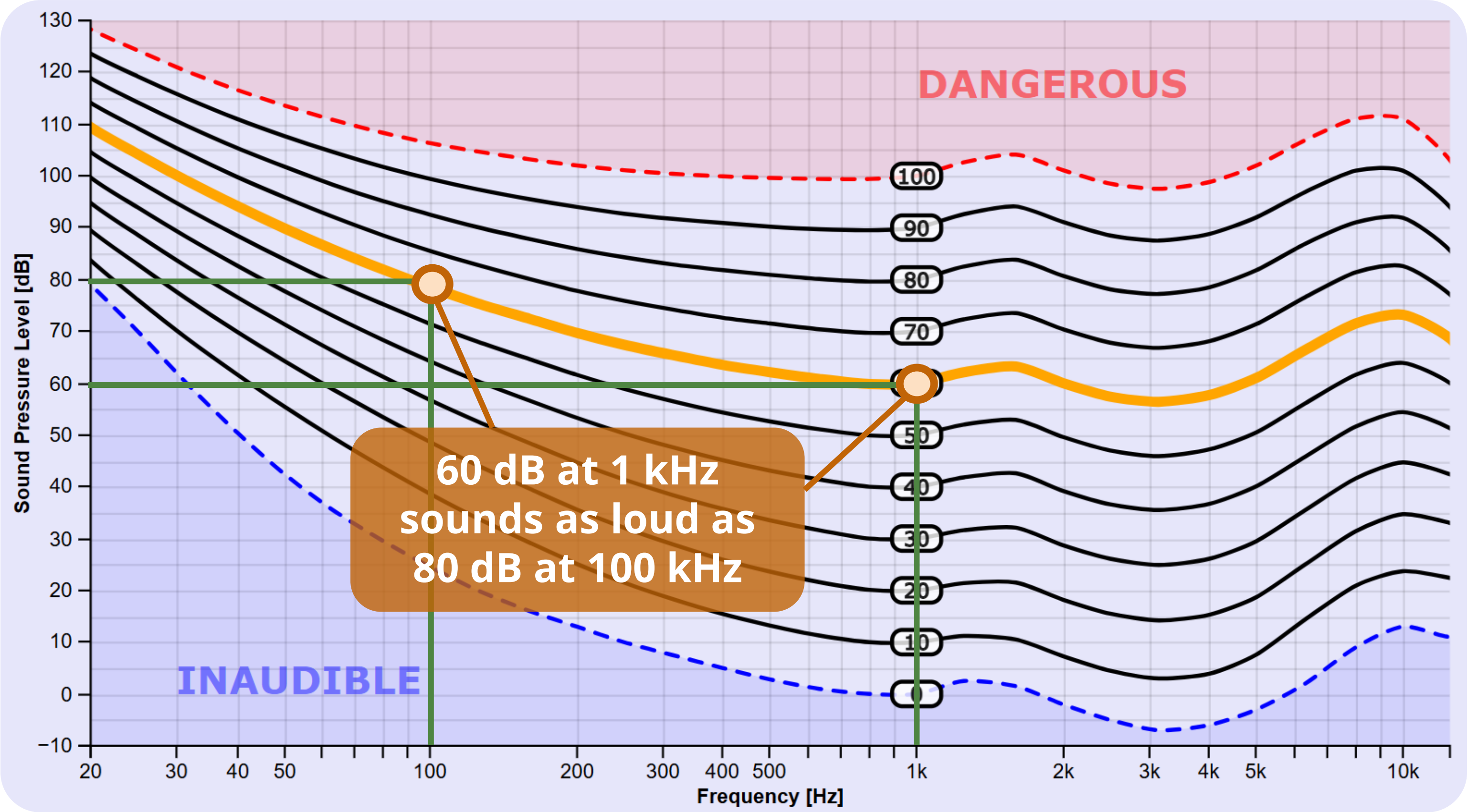
Loudness Perception: Equal-loudness Contours



(Source: Parker, 2024)

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

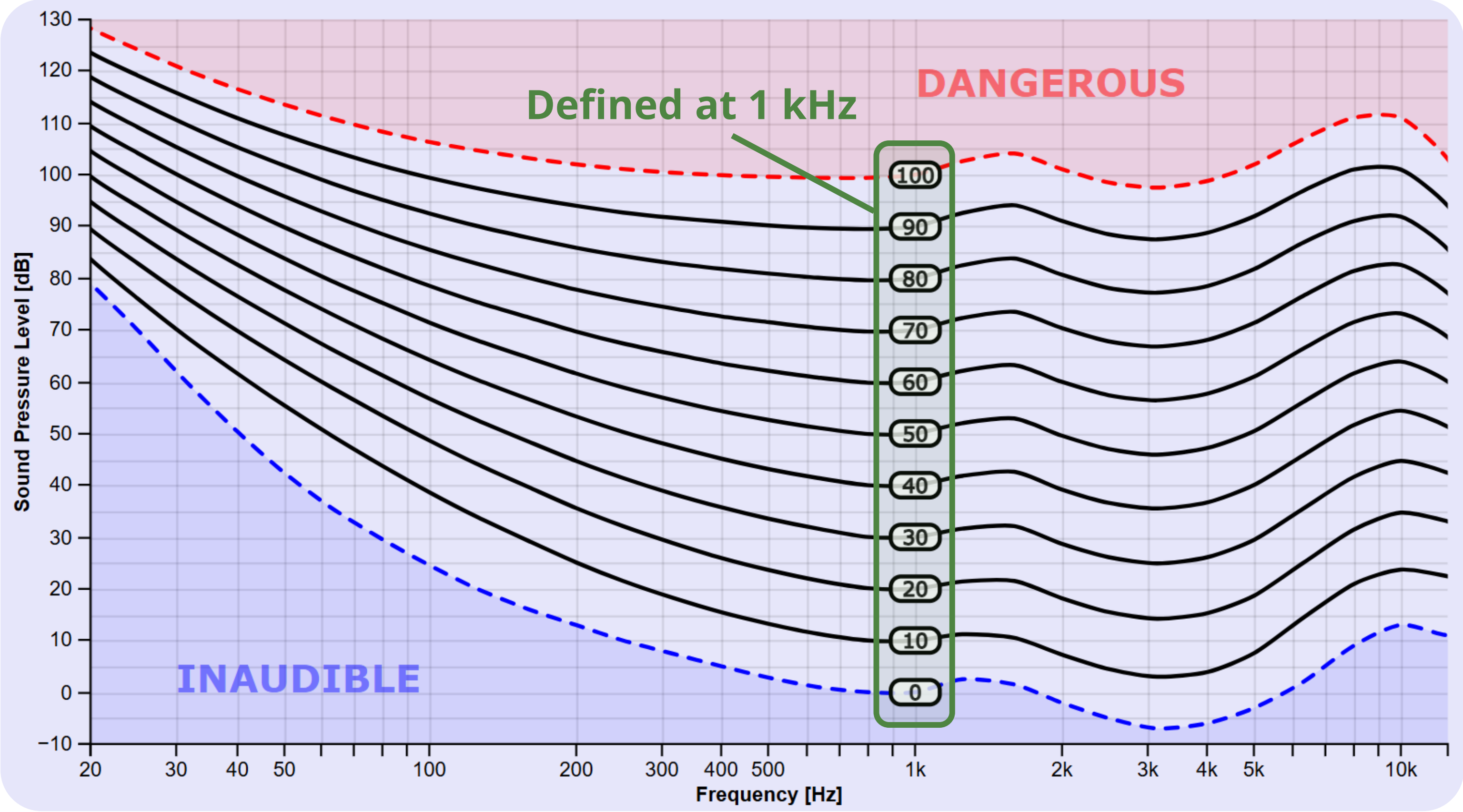
Loudness Perception: Equal-loudness Contours



(Source: Parker, 2024)

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

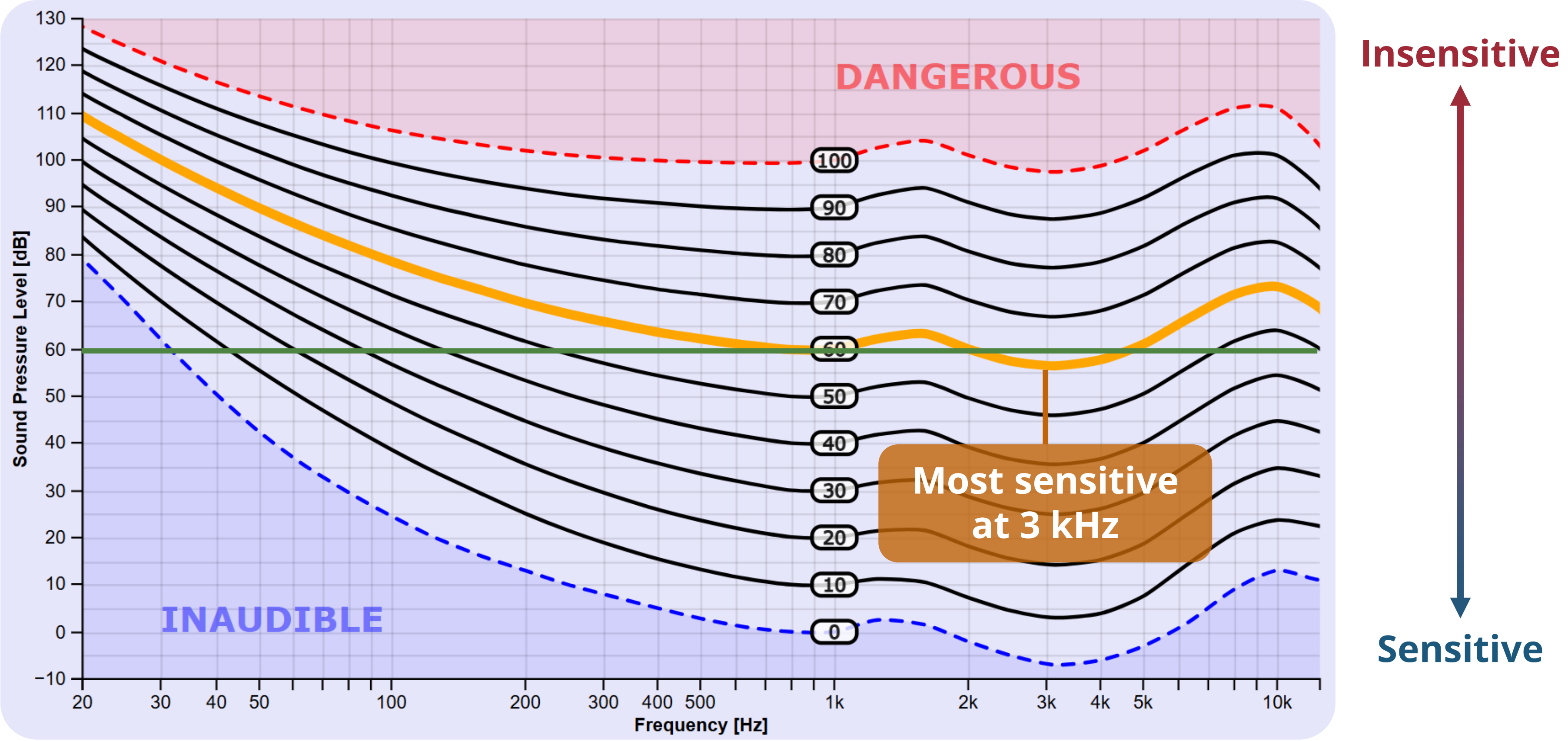
Loudness Perception: Equal-loudness Contours



(Source: Parker, 2024)

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

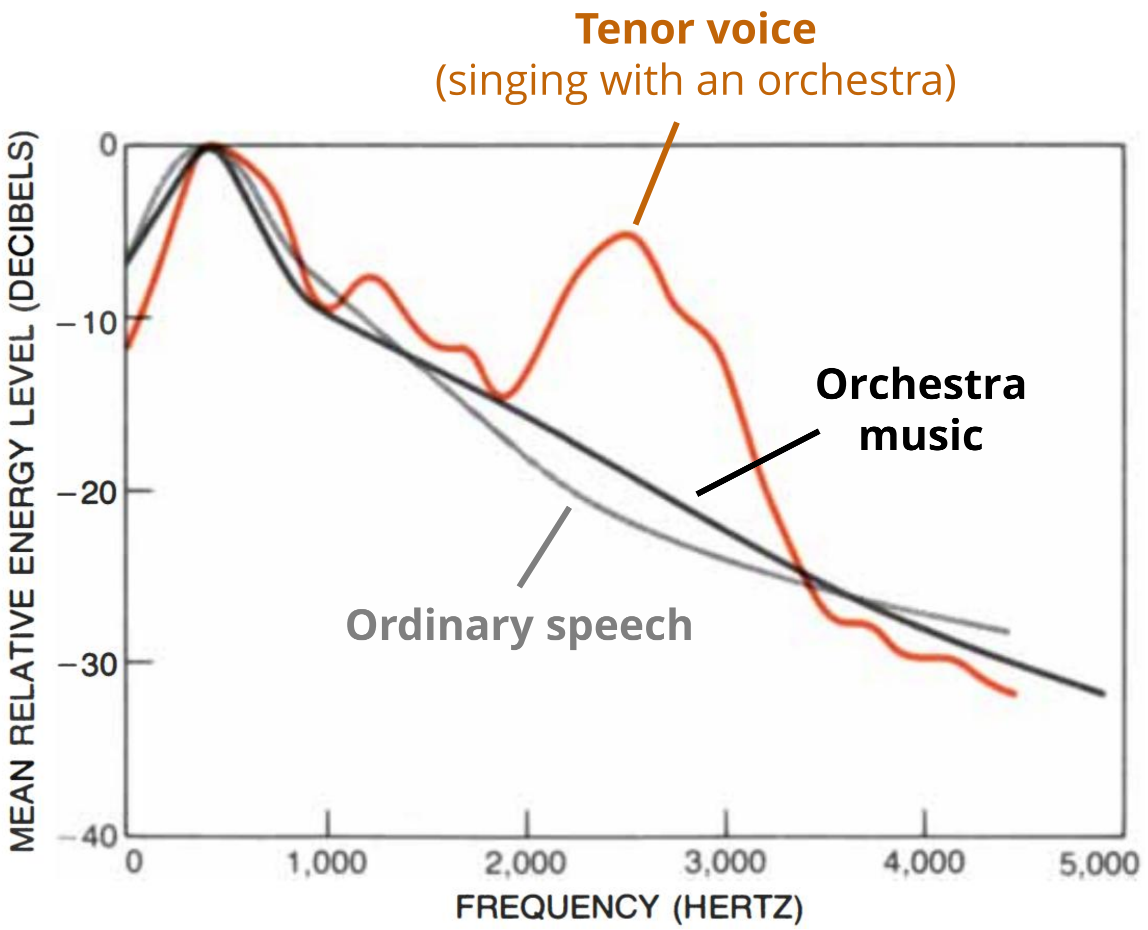
Loudness Perception: Equal-loudness Contours



(Source: Parker, 2024)

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

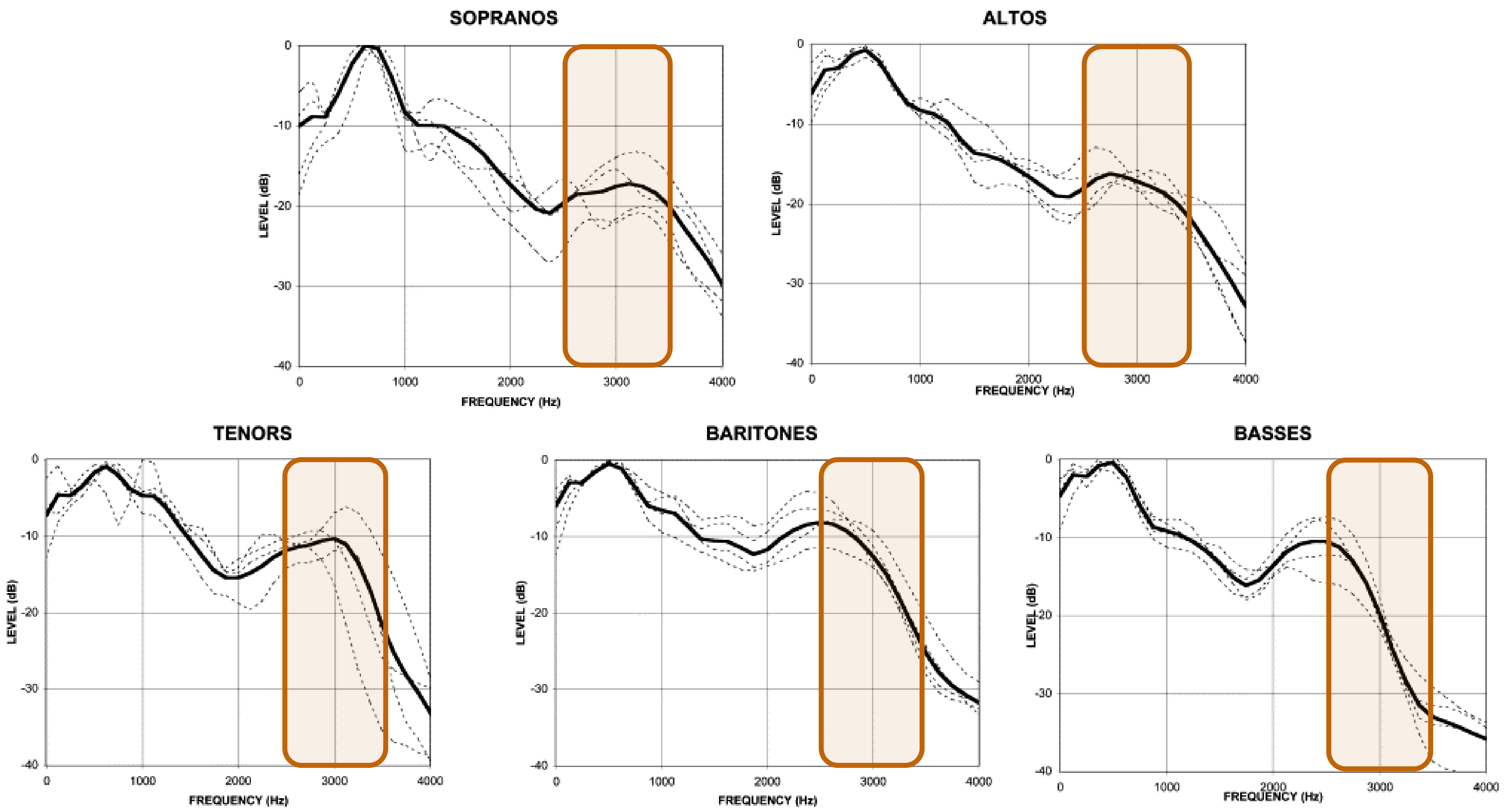
Singer's Formants (Sundberg, 1991)



(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)

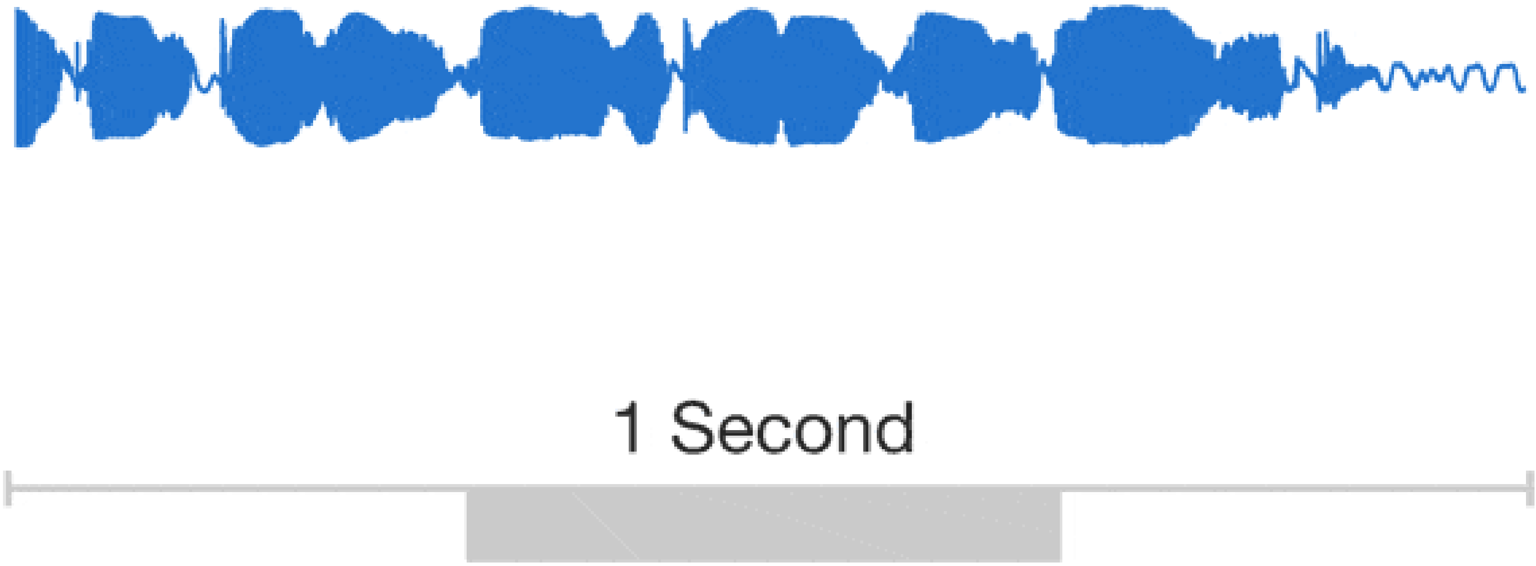
Johan Sundberg, "The Acoustics of the Singing Voice," *Scientific American*, 236(3):82-91, 1977.

Psychoacoustics

- **Acoustics and Psychoacoustics** (PAT 102)
- **Advanced Psychoacoustics** (PAT 421)

Digital Audio

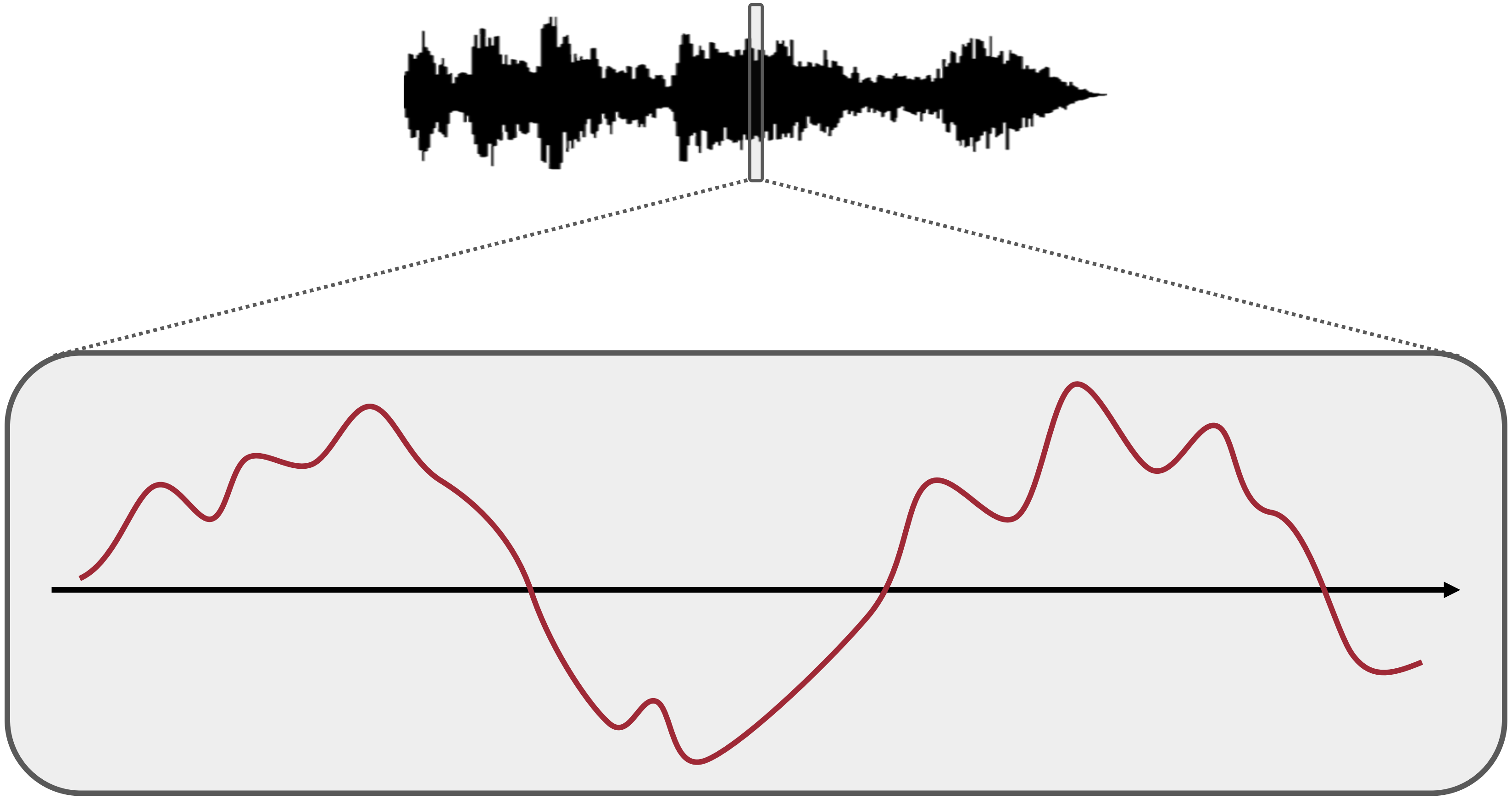
Digital Audio



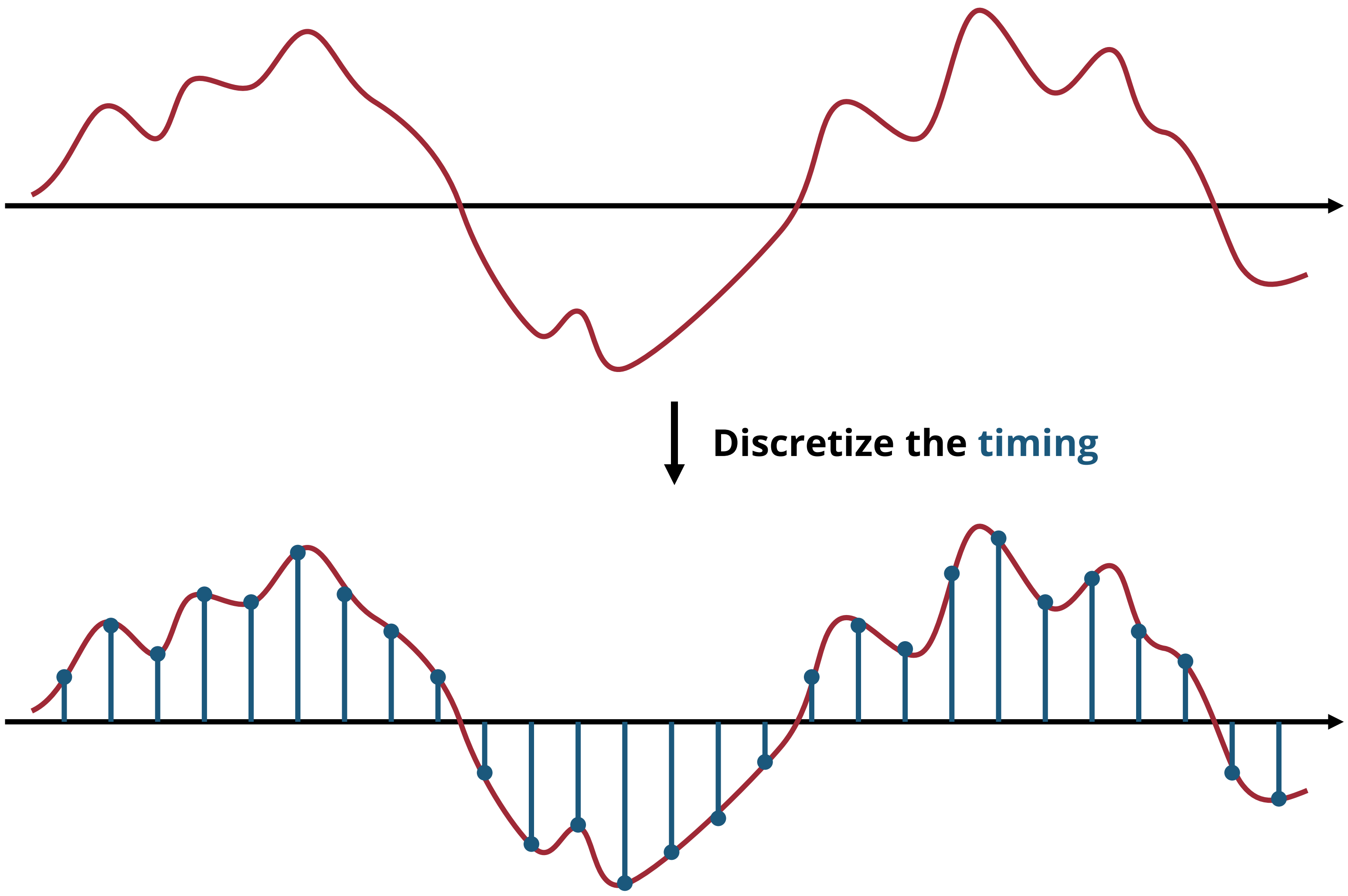
(Source: van den Oord et al., 2016)

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

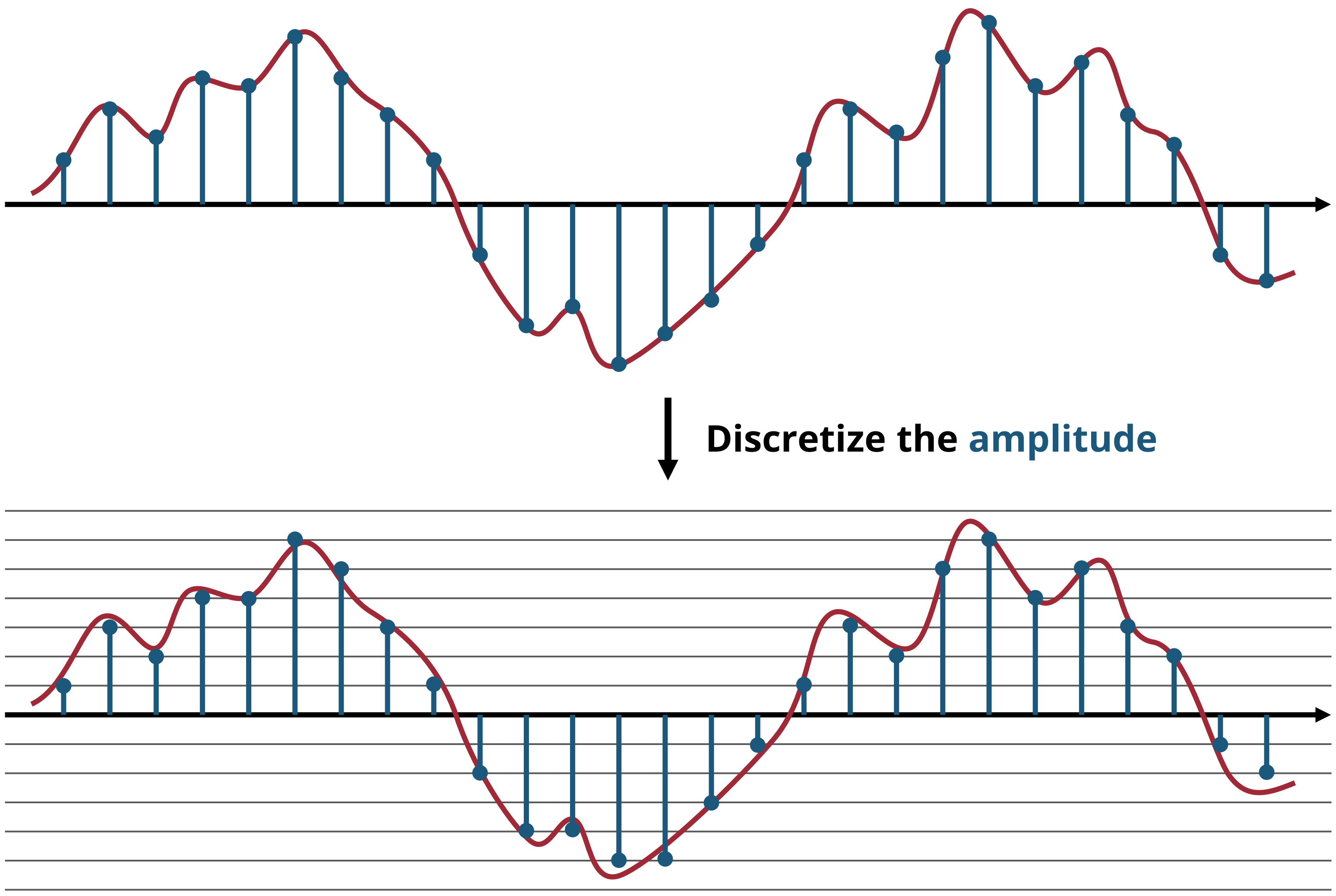
Waveform



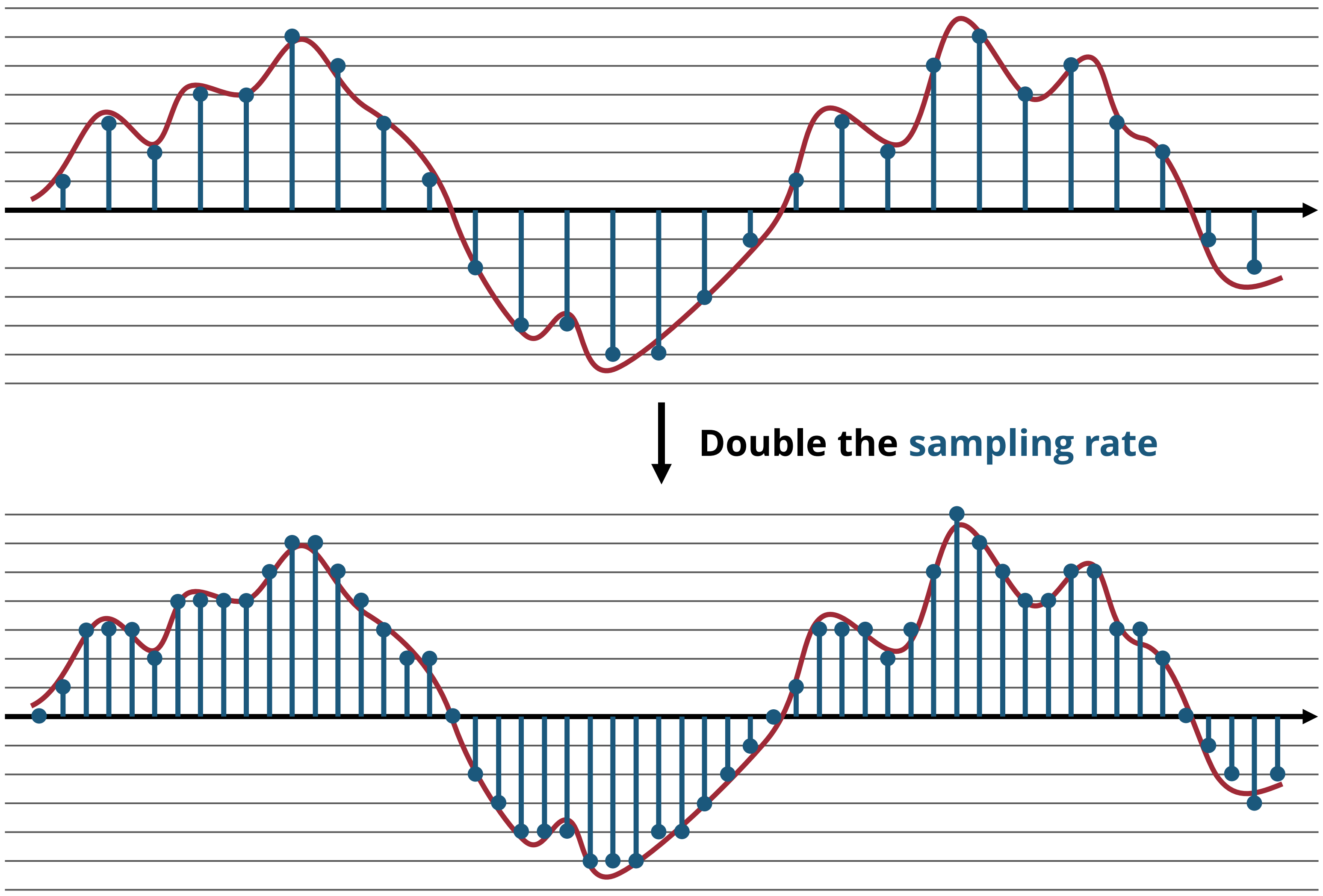
Digitalizing Audio: Timing



Digitalizing Audio: Amplitude



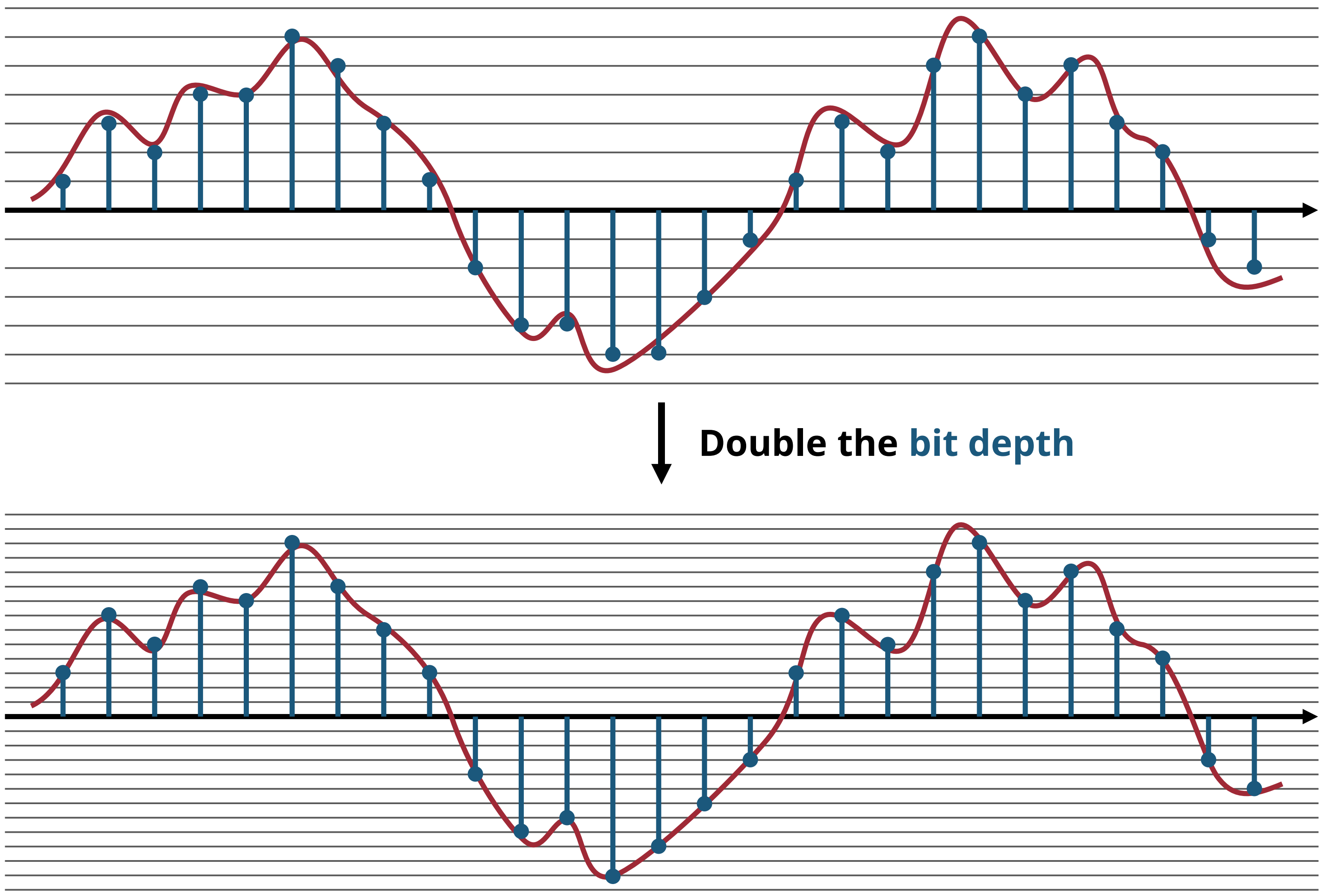
Resolution: Sampling Rate



Sampling Rate

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

Resolution: Bit Depth



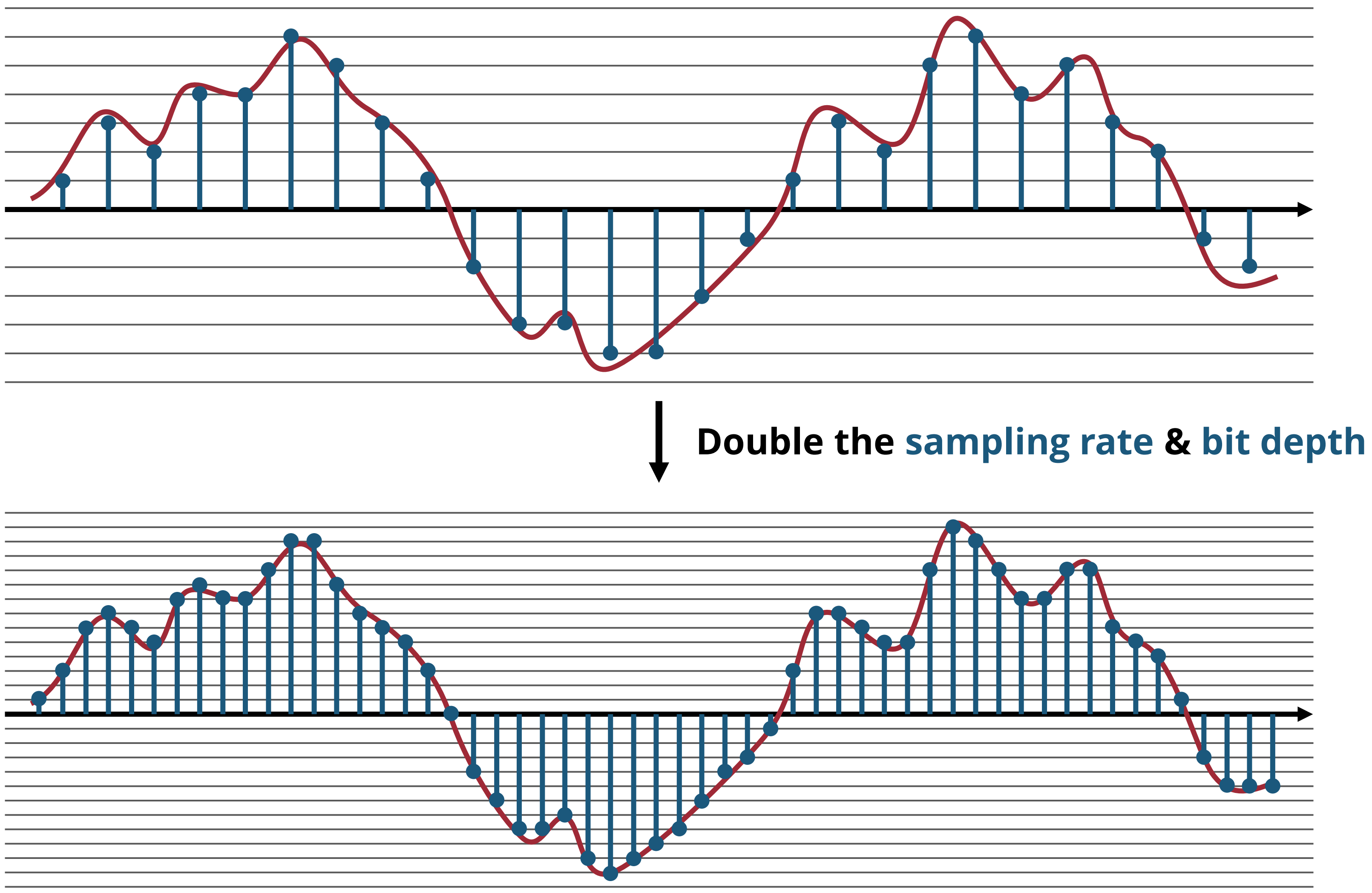
Bit Depth

- **Definition**: **Number of bits used to store each sample**
  - How many bits used to store the amplitude
  - The higher the sampling rate, the lower the 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



Bit Depth ≠ 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!