

PAT 464/564 (Winter 2026)

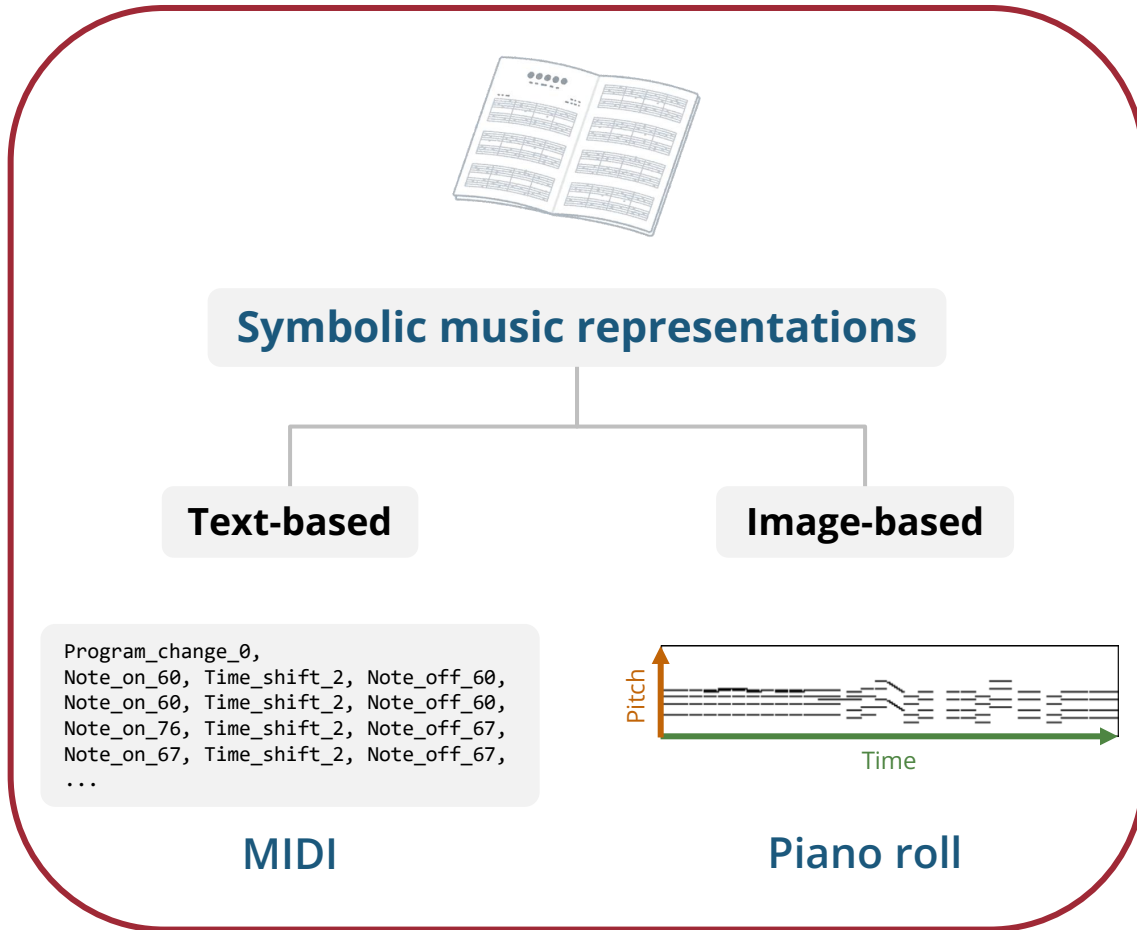
# Generative AI for Music & Audio Creation

## **Lecture 3: Music Processing Fundamentals**

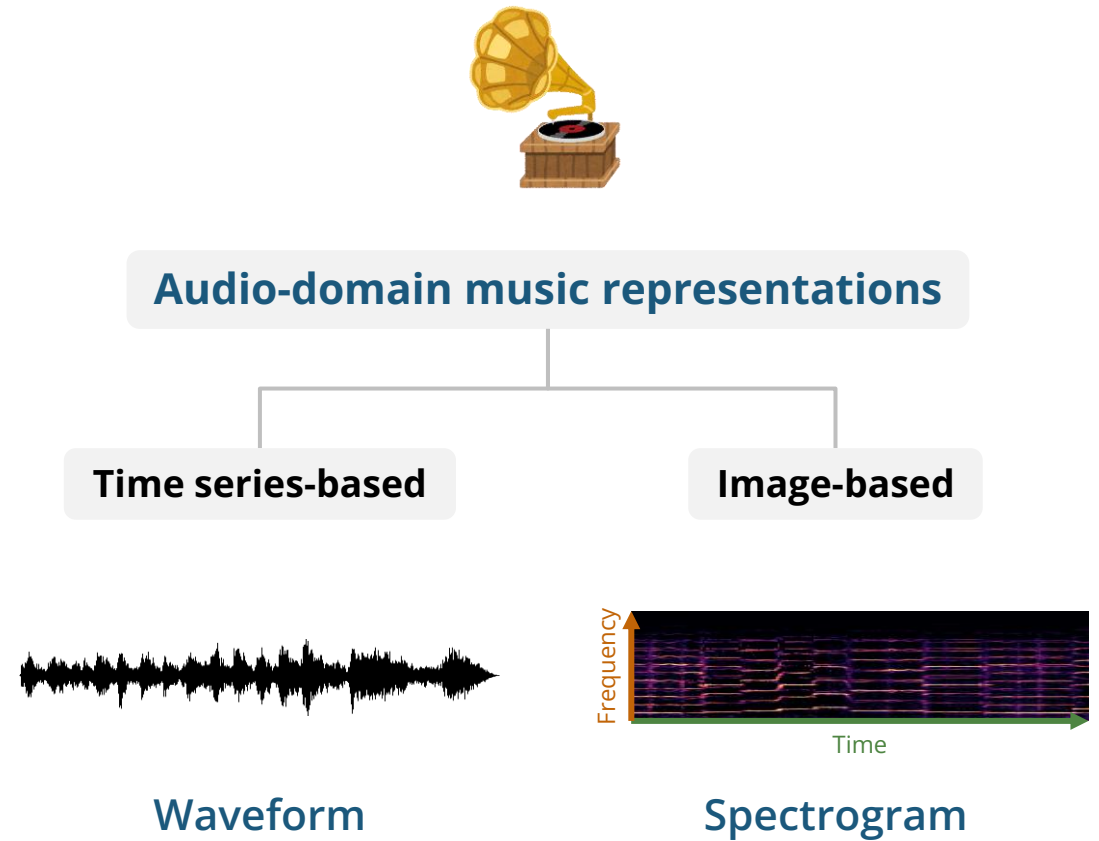
Instructor: Hao-Wen Dong

How do we process music on a computer?

# Four Representative Music Representations



**Today's topic!**



# Machine-readable Music Language

# Designing a Machine-readable Music Language

- How can we represent music in a way that machines understand?



# ABC Notation

- A simple text-based notation
- Use **letters** to denote **itches**
  - Lower octave (A–G), higher octave (a–g)
- Use **prefix** to denote **accidentals**
  - Sharp (^), flat (\_), natural (=)



| 🤔 What is this song in ABC notation?

```
CCGG | AAG2 | FFEE | DDC2 : |  
| : GGFF | EED2 | GGFF | EED2 |  
CCGG | AAG2 | FFEE | DDC2 : |
```

Twinkle, twinkle, little star!

# An Example of ABC Notation

Ah! vous dirai-je, maman  
(Twinkle, twinkle, little star)

*anon. (France)*

♩ = 120

Metadata

```
X:571
T:Ah! vous dirai-je, maman
T:(Twinkle, twinkle, little star)
C:anon.
O:France
R:Nursery song
M:C Meter
L:1/4 Unit note length (temporal resolution)
Q:120 Tempo
K:C Key
CCGG|AAG2|FFEE|DDC2:|
|:GGFF|EED2|GGFF|EED2|
CCGG|AAG2|FFEE|DDC2:|
```

# ChatGPT can Compose Songs in ABC Notation!

**You**  
Write a cheerful folk song using ABC notation

**ChatGPT**  
Certainly! Here's a cheerful folk song in ABC notation:

```
vbnet Copy code  
X:1  
T:Sunshine on the Meadow  
M:4/4  
K:C  
G2 G2 E2 E2 | F2 F2 D4 | E2 E2 G2 G2 | A4 G4 |  
G2 G2 E2 E2 | F2 F2 D4 | E2 E2 G2 G2 | C4 E4 |  
G2 G2 E2 E2 | F2 F2 D4 | E2 E2 G2 G2 | A4 G4 |  
G2 G2 E2 E2 | F2 F2 D4 | G4 F4 | E4 D4 | C4 z4 |  
G2 G2 E2 E2 | F2 F2 D4 | E2 E2 G2 G2 | A4 G4 |  
G2 G2 E2 E2 | F2 F2 D4 | E2 E2 G2 G2 | C4 E4 |  
G2 G2 E2 E2 | F2 F2 D4 | E2 E2 G2 G2 | A4 G4 |  
G2 G2 E2 E2 | F2 F2 D4 | G4 F4 | E4 D4 | C4 z4 |
```

Feel free to play around with the tempo and instrumentation to give it your own unique style!

[chat.openai.com](https://chat.openai.com)

Sunshine on the Meadow



# Large Language Models (LLMs)

- The models behind ChatGPT!

SA

**You**

What's so cool about **AI for music**? Give me a brief answer



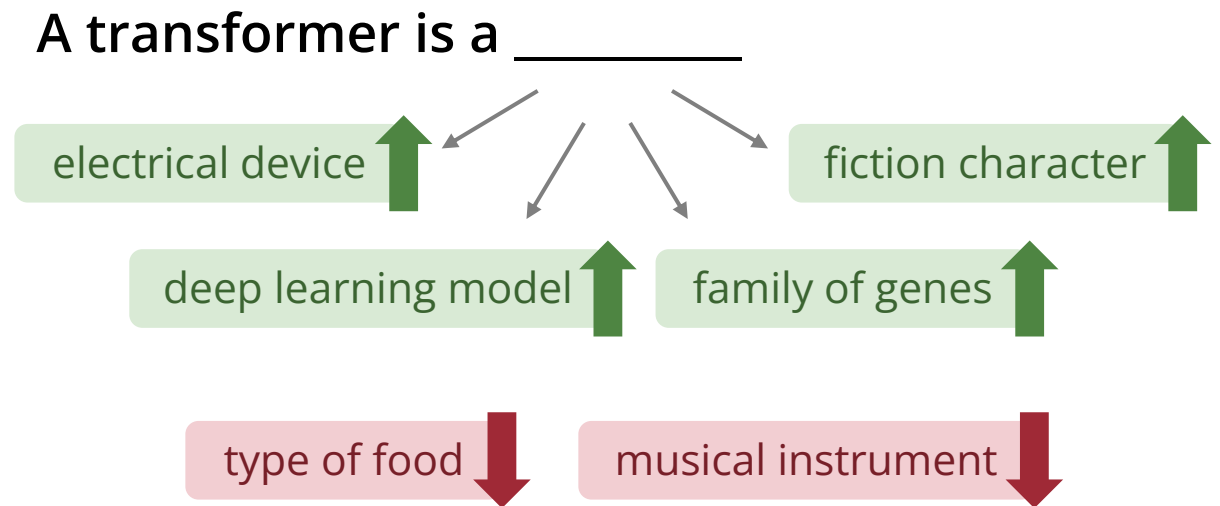
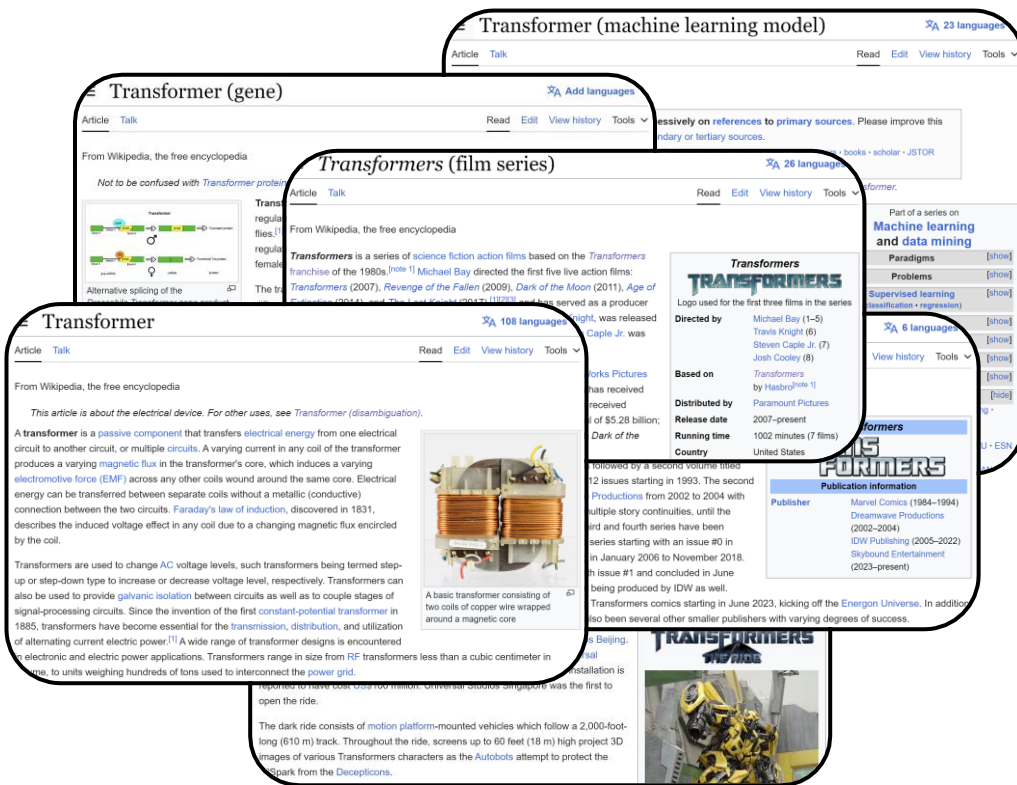
**ChatGPT**

**Word-by-word generation**

AI in music is cool because it can compose original pieces, provide personalized recommendations, automate music production tasks, enhance creativity for artists, enable interactive performances, analyze music trends, and even create virtual artists or bands, expanding the possibilities in music creation and enjoyment.

# Language Models

- Predicting the next word given the past sequence of words



# Language Models (Mathematically)

- A class of machine learning models that **learn** the next word probability

$$P(x_i \mid x_1, x_2, \dots, x_{i-1})$$

Next word      Previous words

$P(\text{electrical} \mid \text{A transformer is a})$  ↑

$P(\text{character} \mid \text{A transformer is a})$  ↑

$P(\text{gene} \mid \text{A transformer is a})$  ↑

$P(\text{model} \mid \text{A transformer is a})$  ↑

$P(\text{food} \mid \text{A transformer is a})$  ↓

$P(\text{musical} \mid \text{A transformer is a})$  ↓

# Music Language Models (Mathematically)

- A class of machine learning models that learn the next note probability

$$P(x_i \mid x_1, x_2, \dots, x_{i-1})$$

Next note      Previous notes

$$\begin{array}{l} P(G \mid C C G G A A) \uparrow \\ P(A \mid C C G G A A) \uparrow \\ P(C \mid C C G G A A) \uparrow \\ P(F \mid C C G G A A) \uparrow \\ P(Ab \mid C C G G A A) \downarrow \\ P(A\# \mid C C G G A A) \downarrow \end{array}$$

# Language Models: Generation

- How do we generate a new sentence using a trained language model?

A transformer is a



A transformer is a deep



A transformer is a deep learning



A transformer is a deep learning model



A transformer is a deep learning model introduced



A transformer is a deep learning model introduced in



# | Limitations of ABC Notations

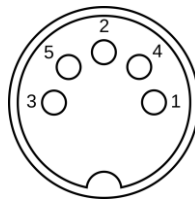
- Limited expressiveness
- Monophonic tunes only

# MIDI-like Representation

# MIDI (Musical Instrument Digital Interface)



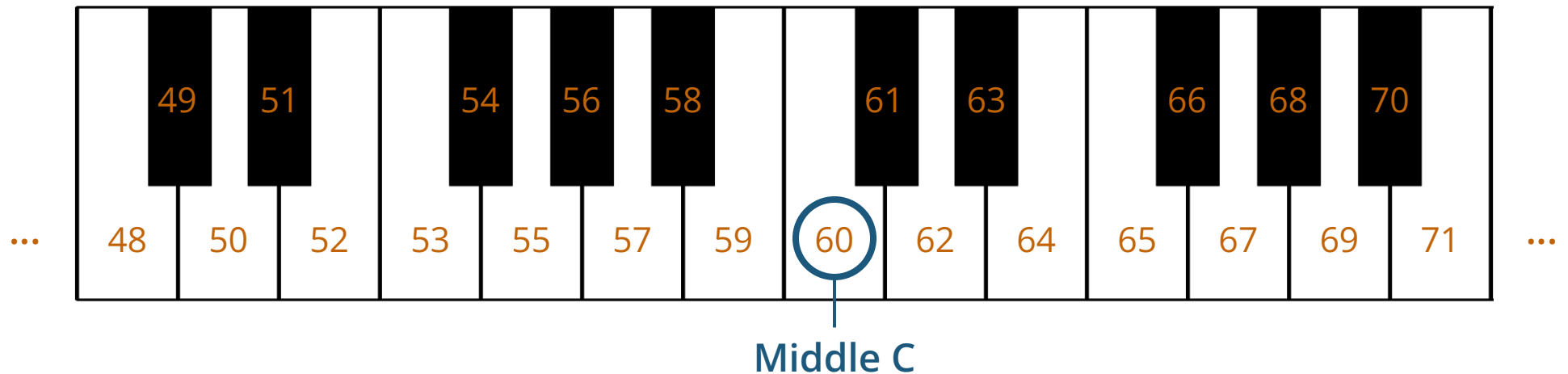
- A communication **protocol** between devices
- MIDI Messages
  - Note on
  - Note off
  - Delta time
  - Program change
  - Control change
  - Pitch bend change



MIDI I/O

# MIDI Note Numbers

- Ranging from 0 to 127
  - Middle C is 60
  - Wider than standard piano's pitch range
- Widely used in various software, keyboards and algorithms



# Representing Music using MIDI Messages

- Three main MIDI messages
  - Note on
  - Note off
  - Time Shift

Sunshine on the Meadow



The image shows two staves of musical notation in 4/4 time. The first staff has a treble clef and a 4/4 time signature. The first two notes are circled: the first is in a blue circle with an orange 'X' over it, and the second is in a green circle. A red arrow points from the first note to the second, indicating a time shift. The second staff continues the melody.

```
Note_on_67 Time_shift_quarter_note Note_off_67  
Note_on_67 Time_shift_quarter_note, Note_off_67,  
Note_on_64, Time_shift_quarter_note, Note_off_64,  
Note_on_64, Time_shift_quarter_note, Note_off_64,  
...
```

# Representing Polyphonic Music

- We can now handle music with multi-pitch at the same time
  - In the literature, “polyphonic” & “multi-pitch” are often used interchangeably

**Clair de Lune**  
from “Suite Bergamasque” L. 75  
3<sup>rd</sup> Movement  
Claude Debussy  
(1862–1918)

*Andante très expressif*

Piano

*pp* *con sordina*

Note\_on\_65, Note\_on\_68, Time\_shift\_eighth\_note, Note\_on\_77, Note\_on\_80,  
Time\_shift\_half\_note, Note\_off\_77, Note\_off\_80, Note\_on\_73, Note\_on\_77,  
Time\_shift\_dotted\_quarter\_note, Note\_off\_65, Note\_off\_68, ...

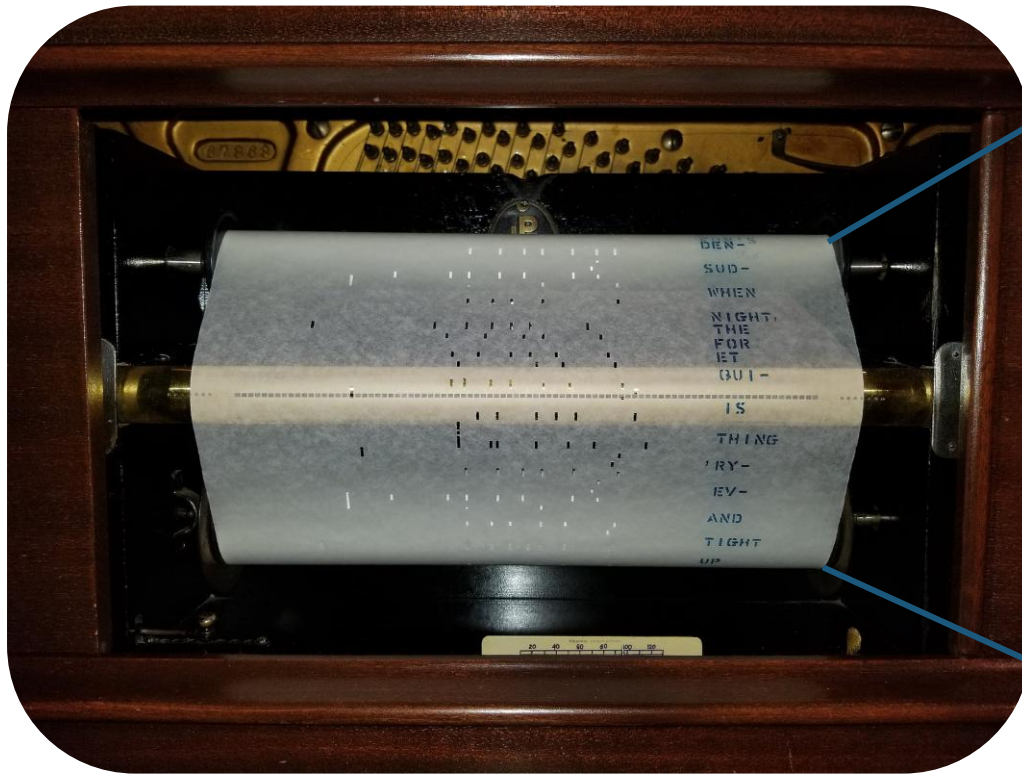
# A.I. Duet (Mann et al, 2016)



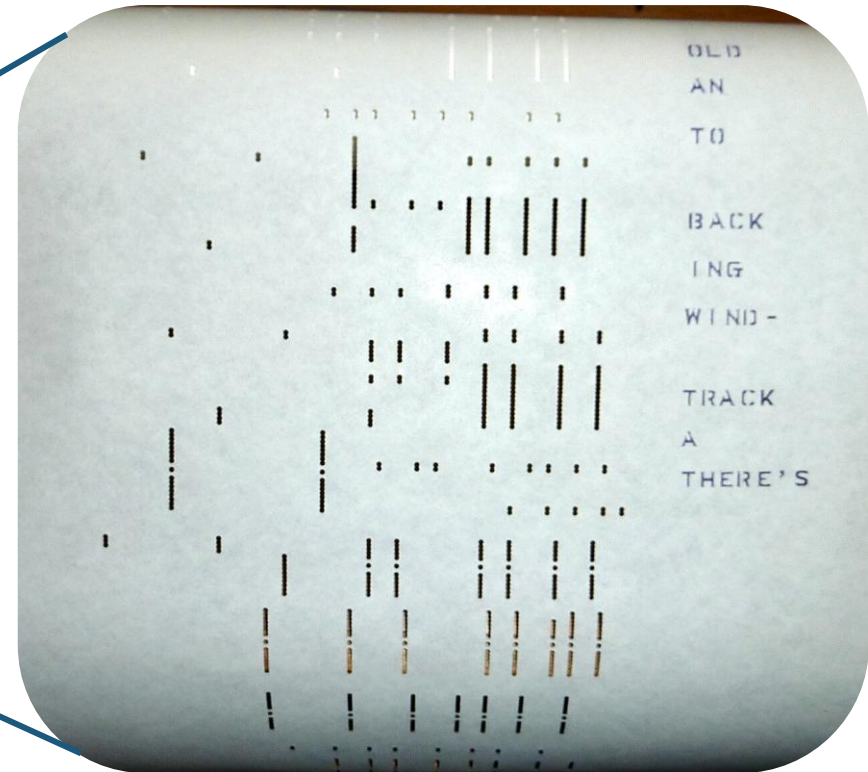
[youtu.be/0ZE1bfPtvZo](https://youtu.be/0ZE1bfPtvZo)  
[experiments.withgoogle.com/ai/ai-duet/view](https://experiments.withgoogle.com/ai/ai-duet/view)

# Piano Roll Representation

# Piano Rolls



(Source: Draconichiaro)



(Source: Tangerineduel)

# | Player Pianos

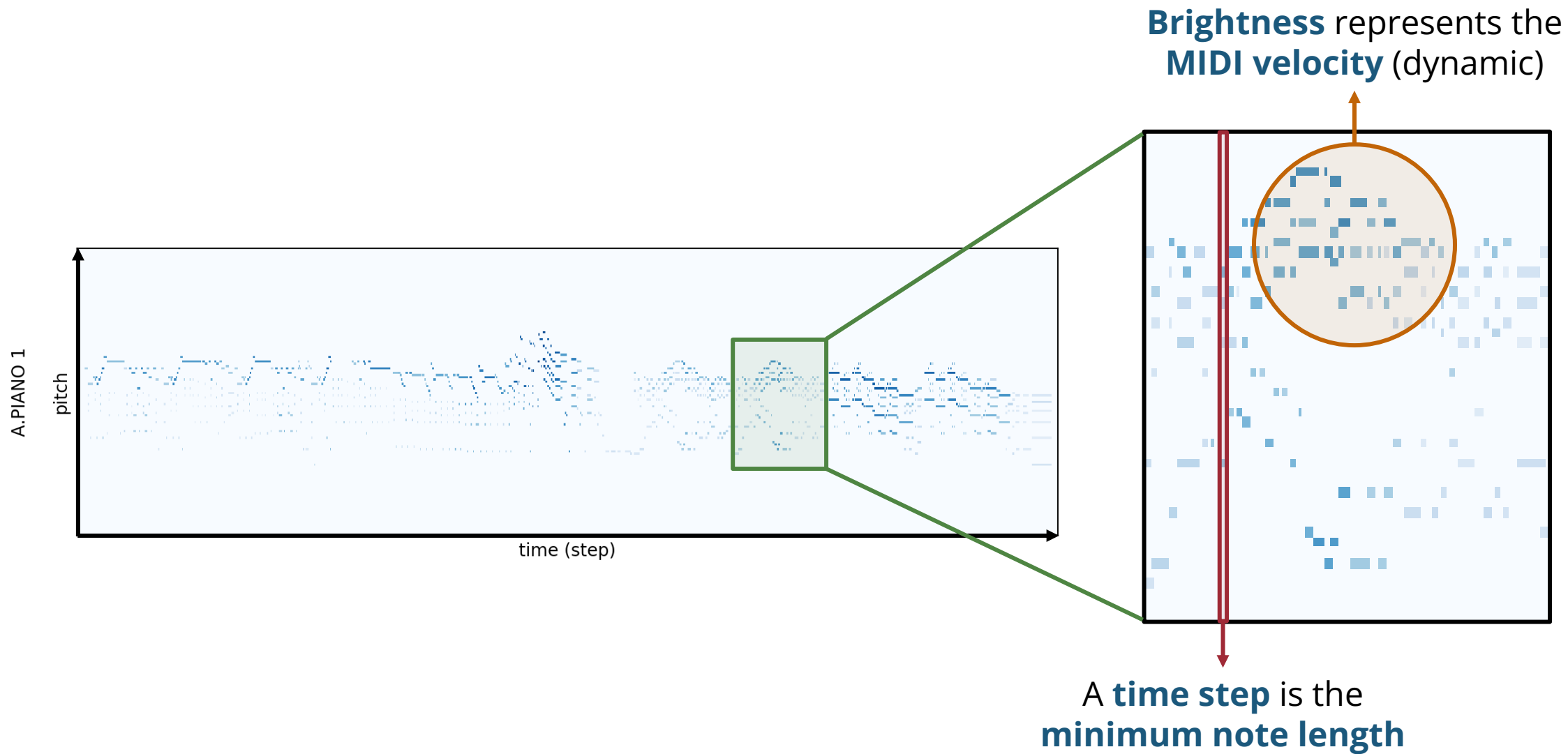


[youtu.be/07krQ661fok](https://youtu.be/07krQ661fok)

# Piano Rolls at the Stearns Collection

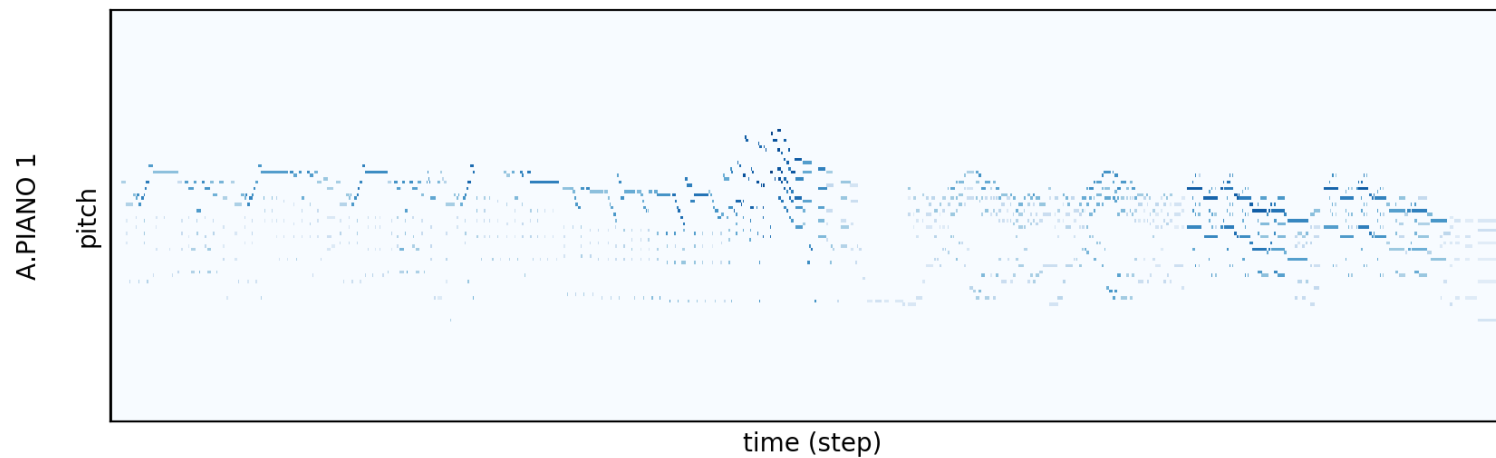


# Piano Roll Representation

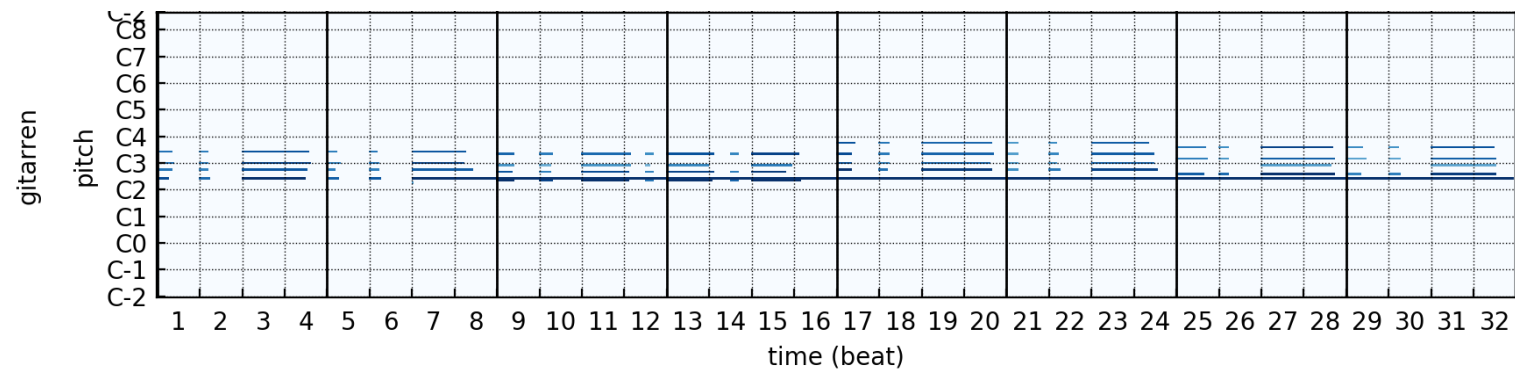


# Piano Roll Representation

With expressive timing

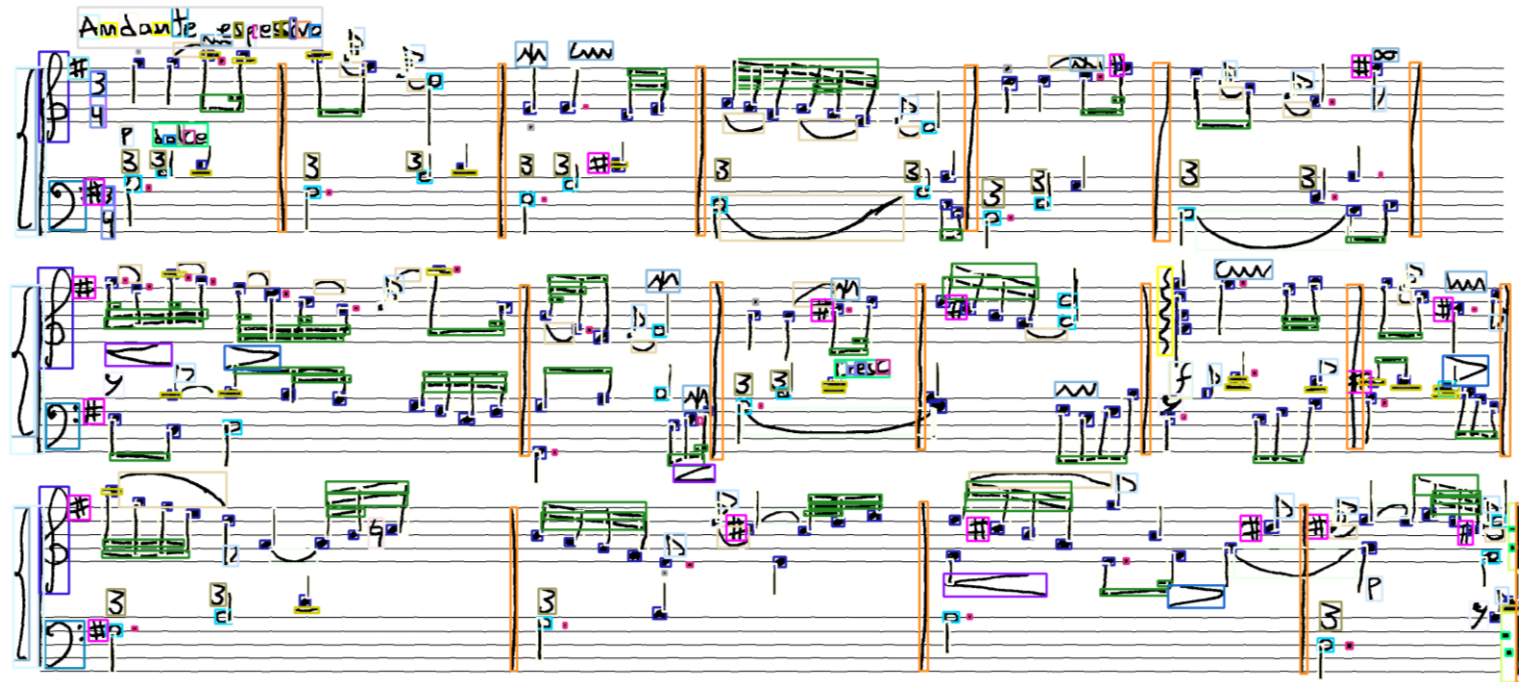


Without expressive timing



# 🤔 Can't we use sheet music "scans" directly?

- Machines still have a hard time reading scanned sheet music
- A task known as **optical music recognition (OMR)**

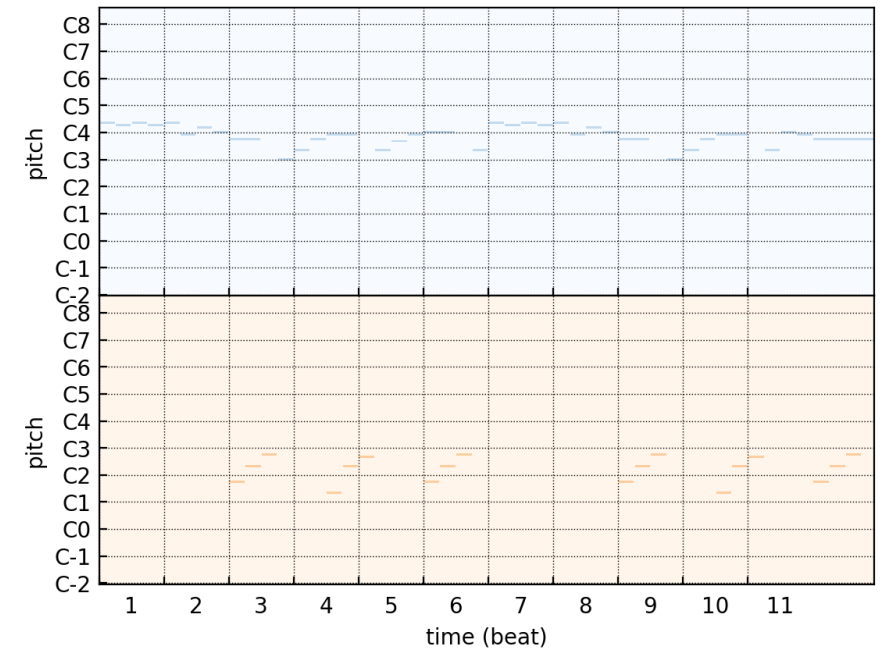


(Source: Pacha et al.)

# Processing Symbolic Music in Python

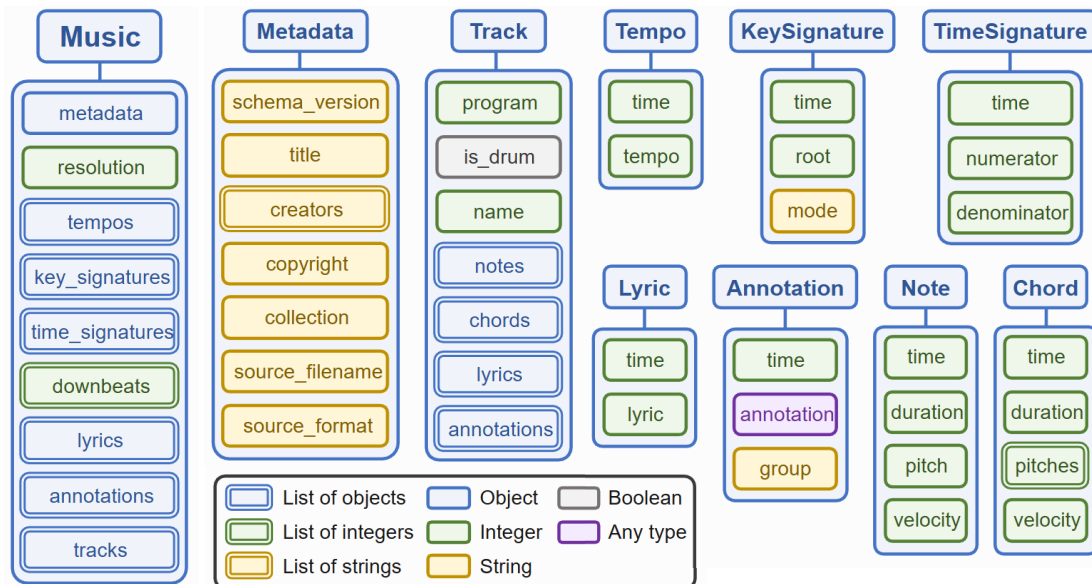
# 🔥 PA1: Symbolic Music Processing

- Use [MusPy](#) to process symbolic music
  - **Read** MIDI and MusicXML files
  - **Manipulate** notes, instruments, and tempos
  - **Show** as piano roll(s)
  - **Write** as MIDI and MusicXML files
  - **Synthesize** into WAV files
- Learn to code in [Jupyter Notebook](#)
- Learn about [Google Colab](#)



# muspy.Music Class

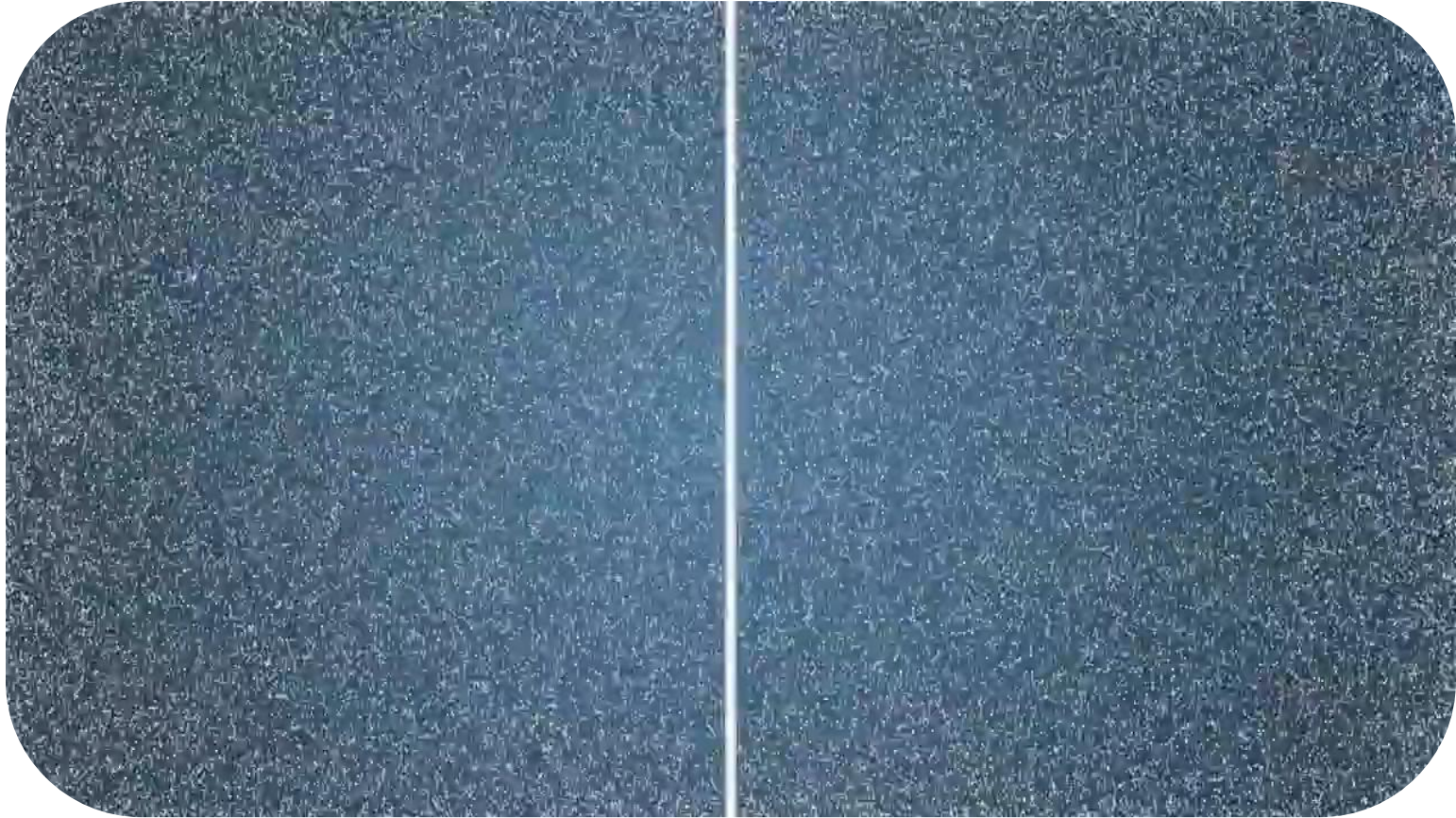
- Core class of MusPy
- A universal container for symbolic music
- Serializable to JSON/YAML



```
metadata:
  schema_version: '0.0'
  title: Für Elise
  creators: [Ludwig van Beethoven]
  collection: Example dataset
  source_filename: example.json
resolution: 4
tempos:
  - {time: 0, qpm: 72.0}
key_signatures:
  - {time: 0, root: 9, mode: minor}
time_signatures:
  - {time: 0, numerator: 3, denominator: 8}
downbeats: [4, 16]
lyrics:
  - {time: 0, lyric: Nothing but a lyric}
annotations:
  - {time: 0, annotation: Nothing but an annotation}
tracks:
  - program: 0
    is_drum: false
    name: Melody
    notes:
      - {time: 0, duration: 2, pitch: 76, velocity: 64}
      - {time: 2, duration: 2, pitch: 75, velocity: 64}
      - {time: 4, duration: 2, pitch: 76, velocity: 64}
      - {time: 6, duration: 2, pitch: 75, velocity: 64}
      - {time: 8, duration: 2, pitch: 76, velocity: 64}
      - {time: 10, duration: 2, pitch: 71, velocity: 64}
      - {time: 12, duration: 2, pitch: 74, velocity: 64}
      - {time: 14, duration: 2, pitch: 72, velocity: 64}
      - {time: 16, duration: 2, pitch: 69, velocity: 64}
    lyrics:
      - {time: 0, lyric: Nothing but a lyric}
    annotations:
      - {time: 0, annotation: Nothing but an annotation}
```

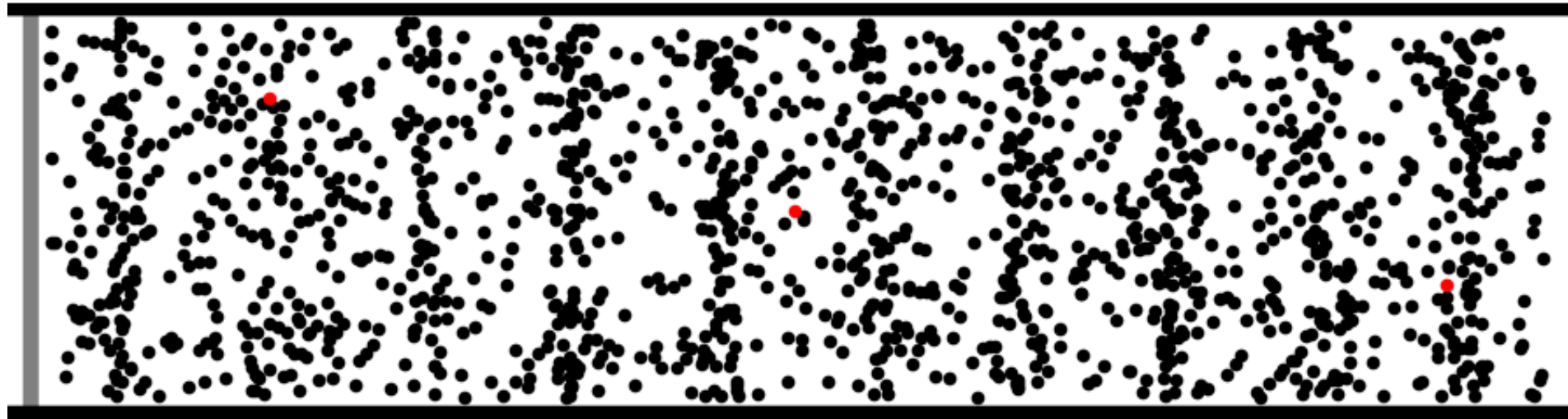
# What is Sound?

# | What is Sound?



[youtu.be/aPswnDcteS4](https://youtu.be/aPswnDcteS4)

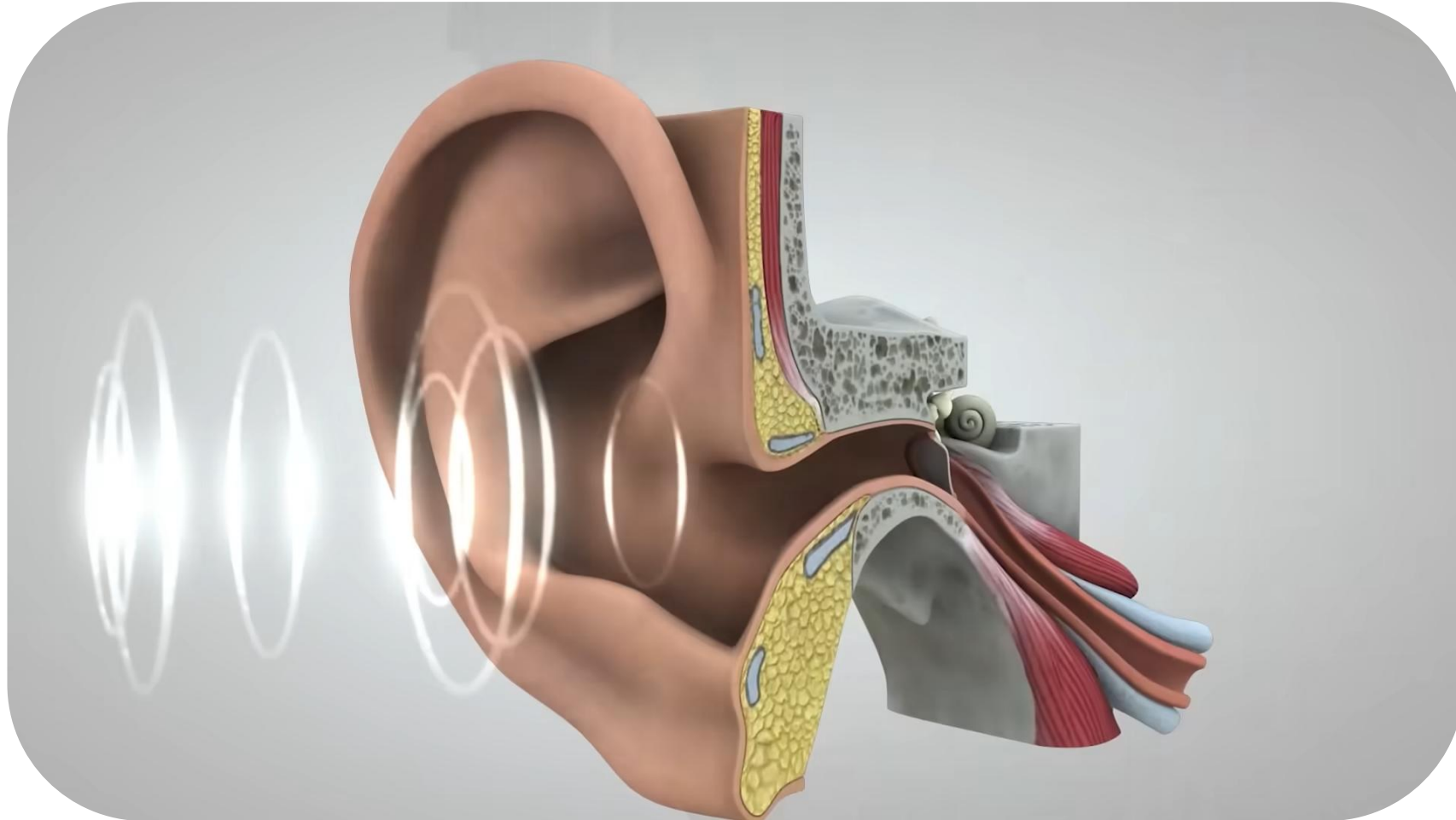
# Sound is a Pressure Wave



©2011. Dan Russell

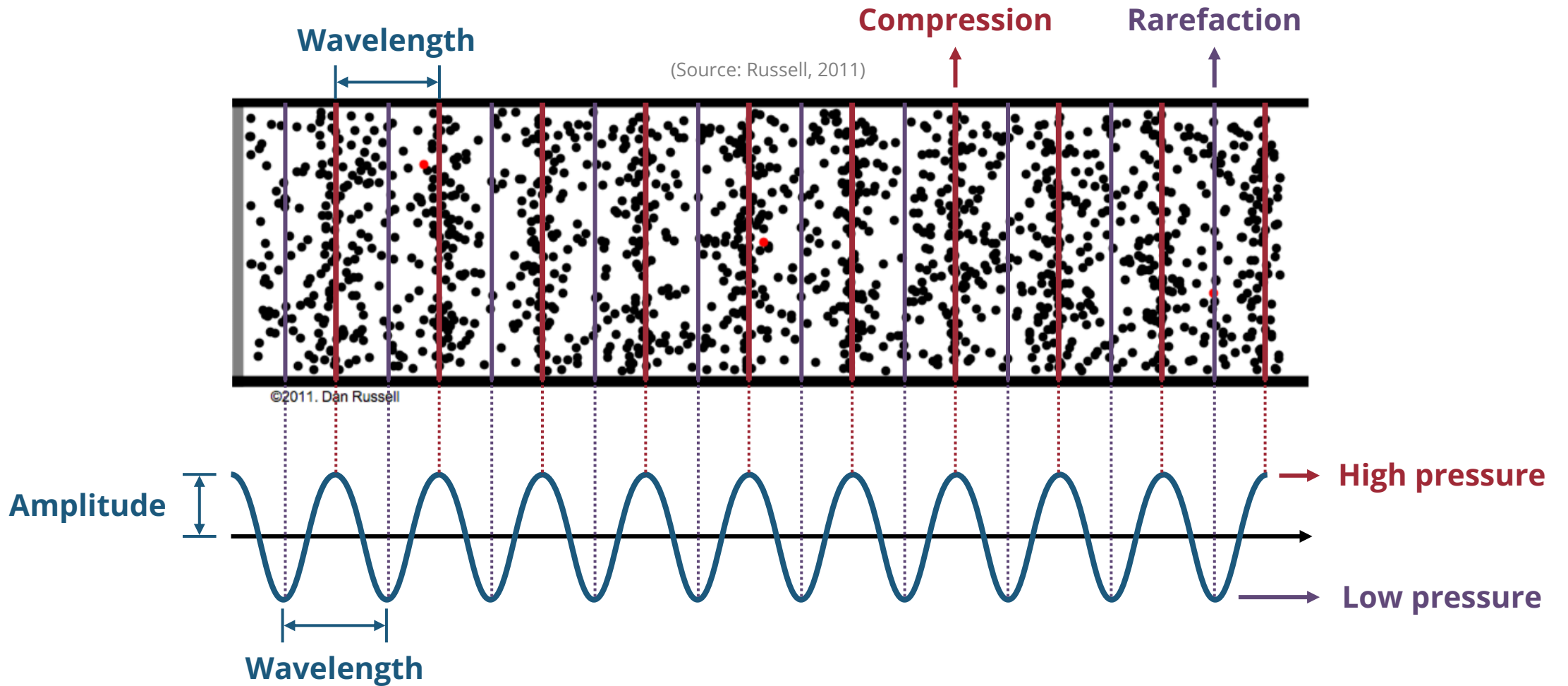
(Source: Russell, 2011)

# Human Auditory System



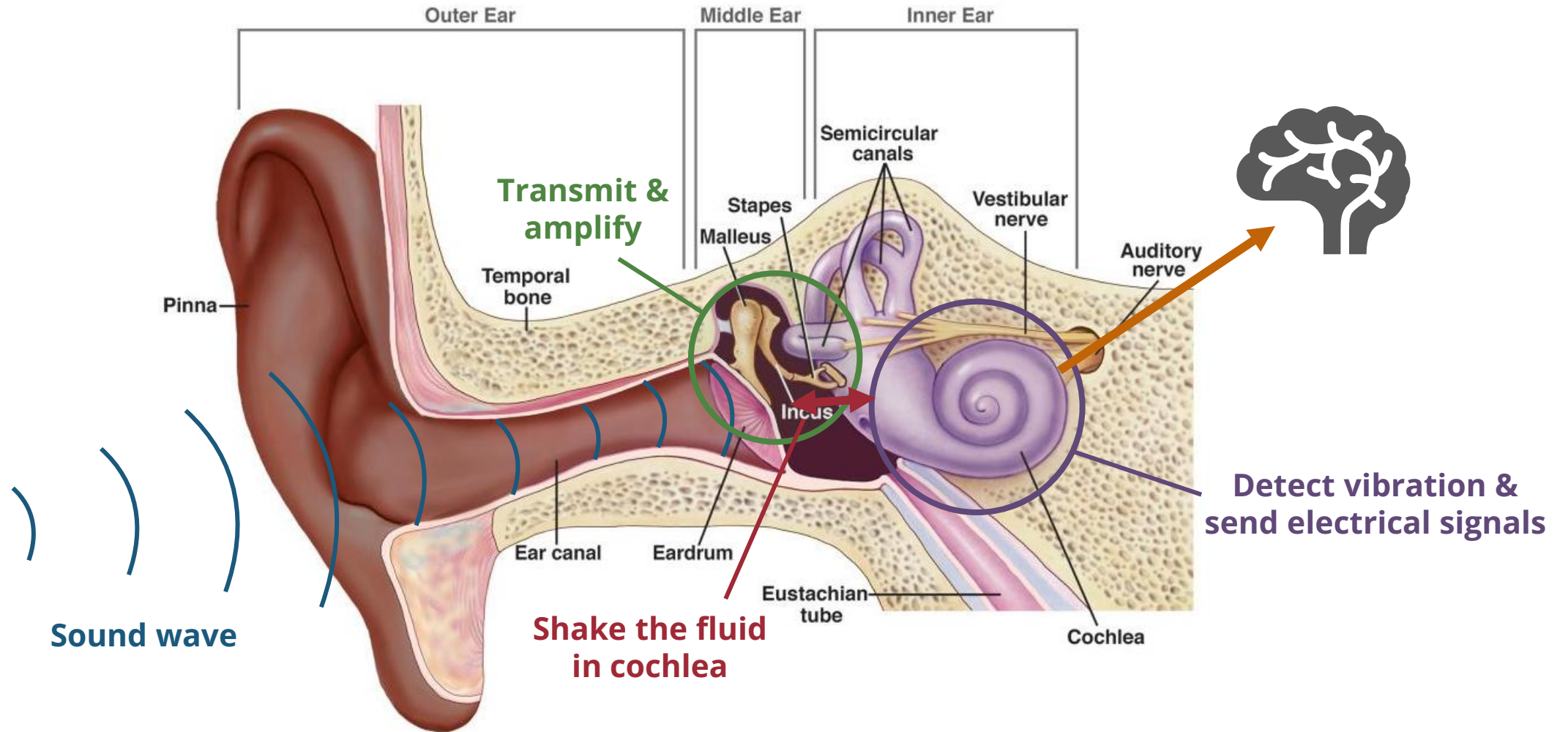
[youtu.be/eQEaiZ2j9oc](https://youtu.be/eQEaiZ2j9oc)

# Longitudinal vs Transverse Waves



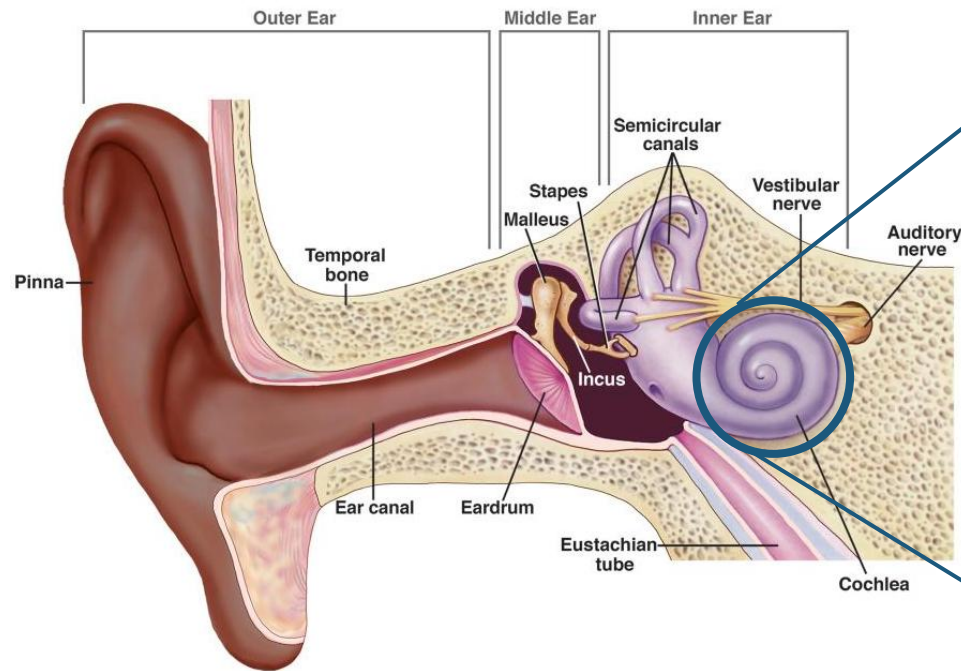
# Auditory Perception

# Human Ears

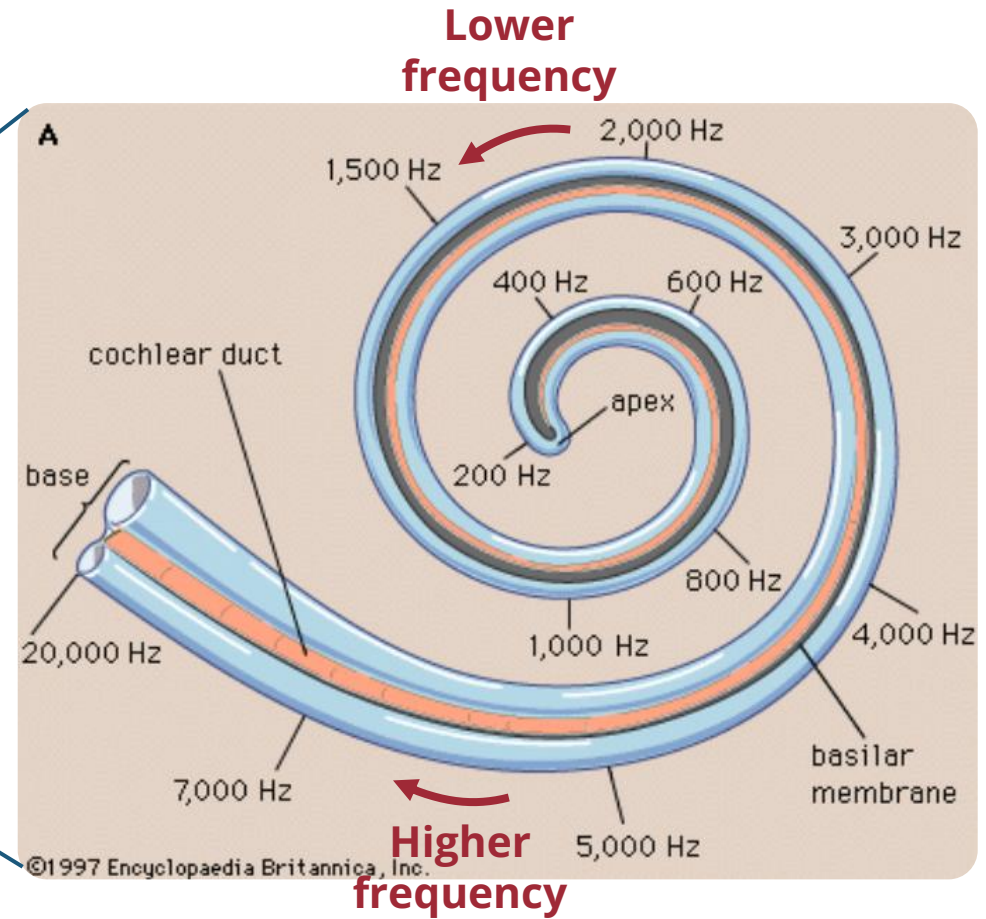


(Source: NIH/NIDCD)

# Cochlea in the Inner Ear



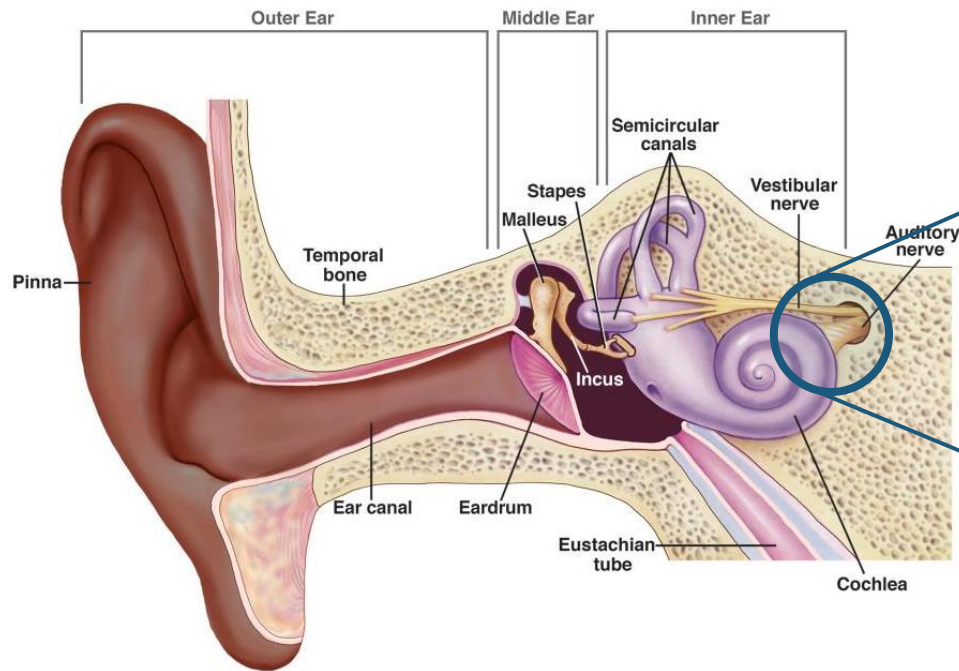
(Source: NIH/NIDCD)



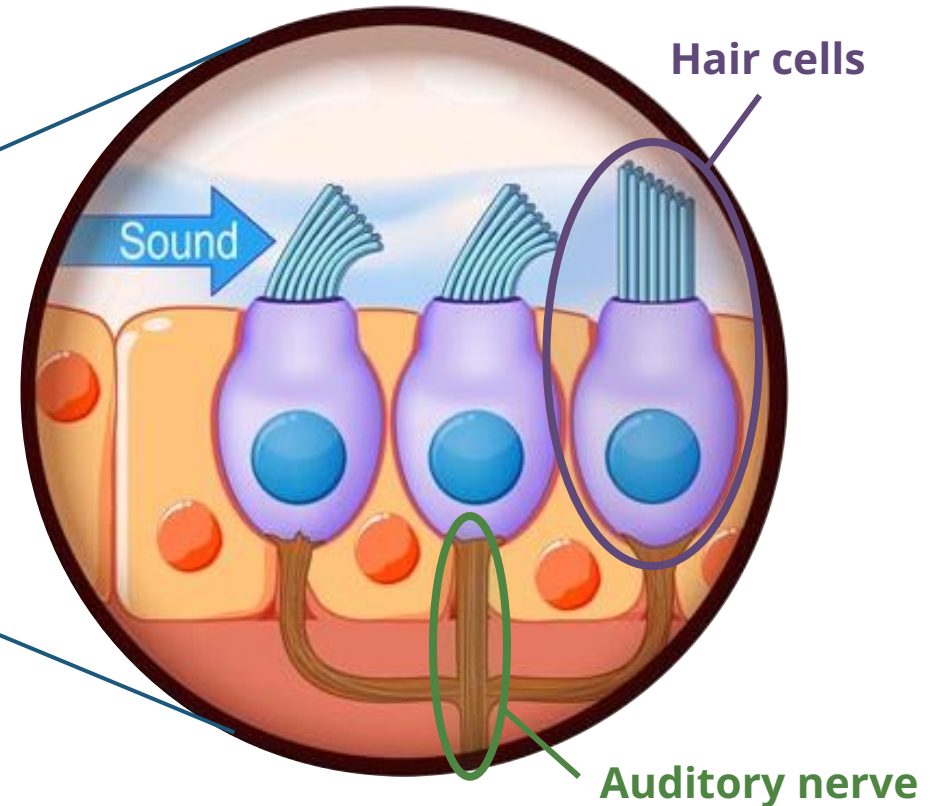
(Source: Britannica)

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



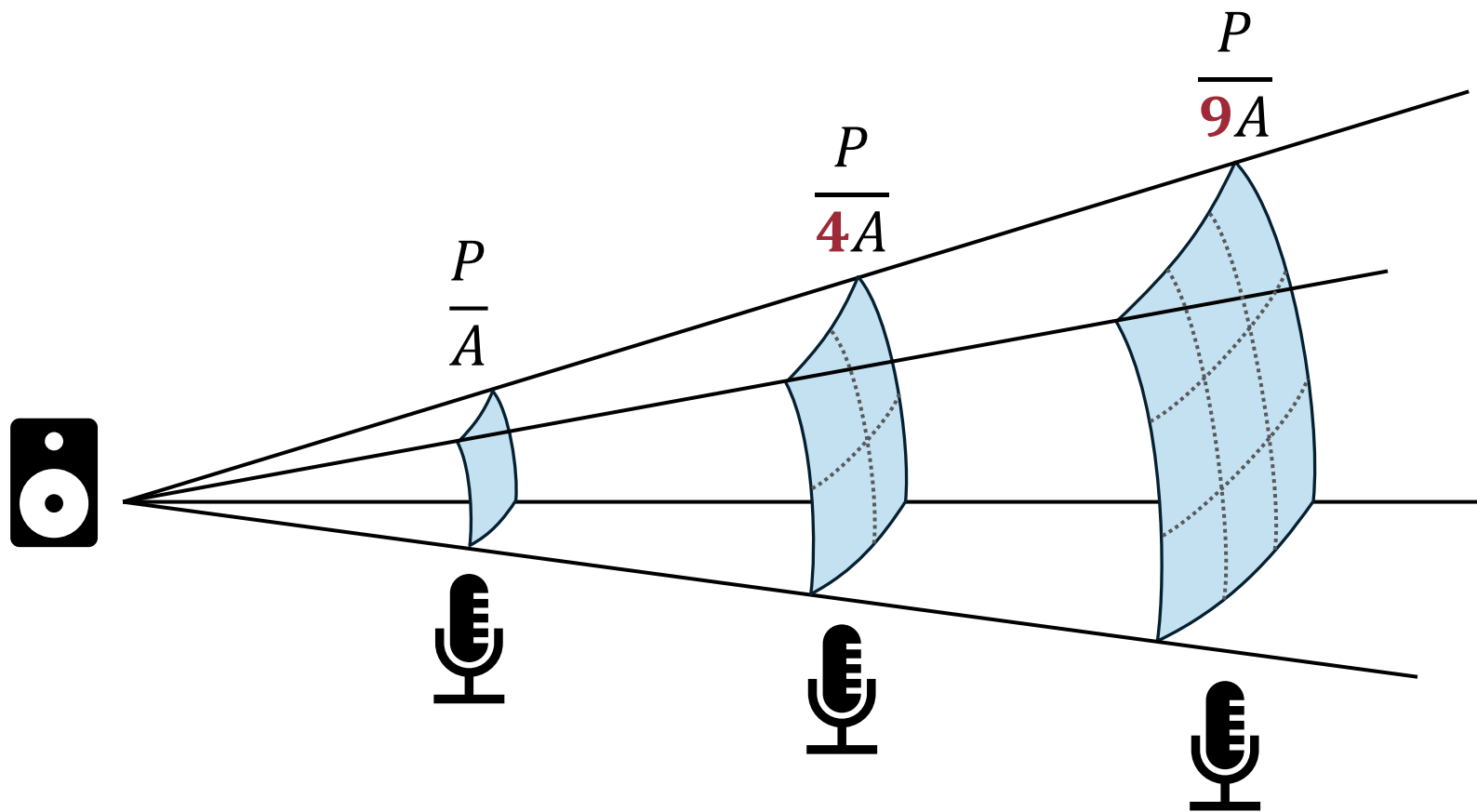
(Source: NIH/NIDCD)



(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 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

$$I_{dB} := 10 \log_{10} \left( \frac{I}{I_{REF}} \right)$$

- $I_{REF} := 10^{-12} W/m^2$  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	
	<b>130</b>	10	Artillery fire at close proximity ( <b>threshold of pain</b> )	
	<b>120</b>	1	Amplified rock music; near jet engine	
	<b>110</b>	$10^{-1}$	Loud orchestral music, in audience	
	<b>100</b>	$10^{-2}$	Electric saw	
	<b>90</b>	$10^{-3}$	Bus or truck interior	
	<b>80</b>	$10^{-4}$	Automobile interior	
	<b>70</b>	$10^{-5}$	Average street noise; loud telephone bell	
	<b>60</b>	$10^{-6}$	<b>Normal conversation</b> ; business office	
	<b>50</b>	$10^{-7}$	Restaurant; private office	
	<b>40</b>	$10^{-8}$	Quiet room in home	
	<b>30</b>	$10^{-9}$	Quiet lecture hall; bedroom	
	<b>20</b>	$10^{-10}$	Radio, television, or recording studio	
	<b>10</b>	$10^{-11}$	Soundproof room	
	<b>0</b>	$10^{-12}$	Absolute silence ( <b>threshold of hearing</b> )	

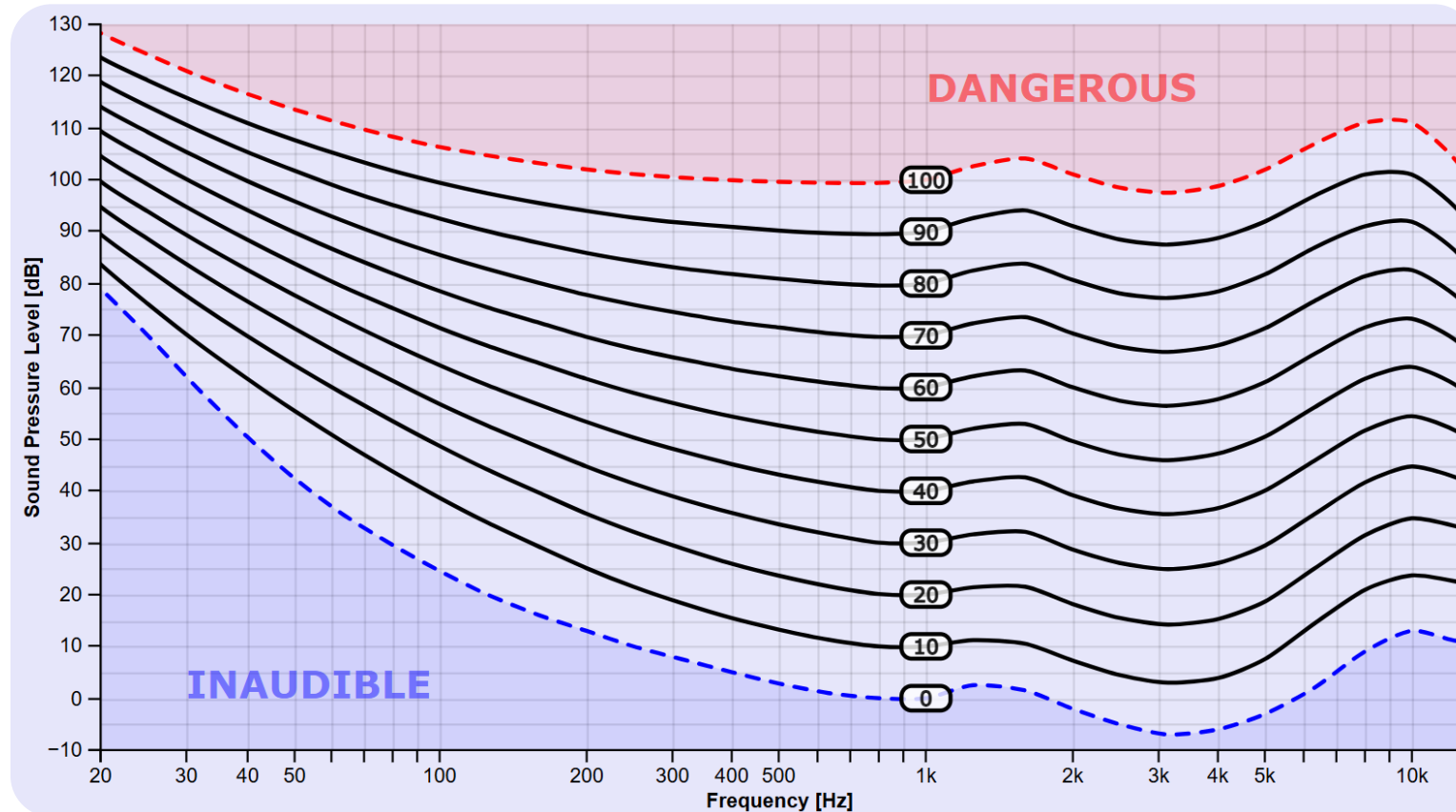
(Unit: W/m<sup>2</sup>)

(Source: Britannica)

## 🤔 Common Gains in Decibels

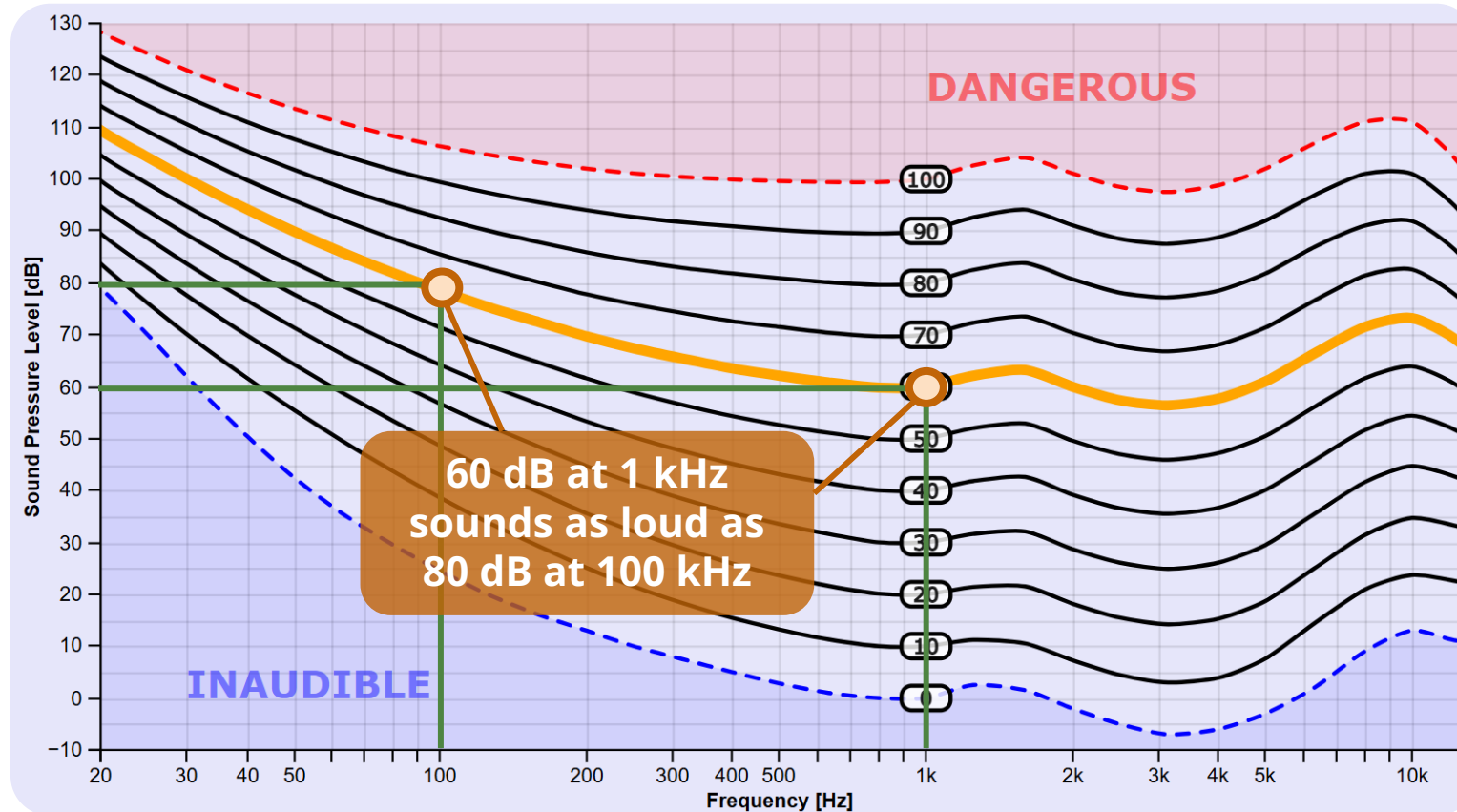
- **+10 dB = 10x intensity  $\approx$  3.2x amplitude**
- **+3 dB  $\approx$  2x intensity  $\approx$  1.4x amplitude**
  
- **2x amplitude = 4x intensity  $\approx$  +6 dB**
- **10x amplitude = 100x intensity = +20 dB**

# Loudness Perception: Equal-loudness Contours



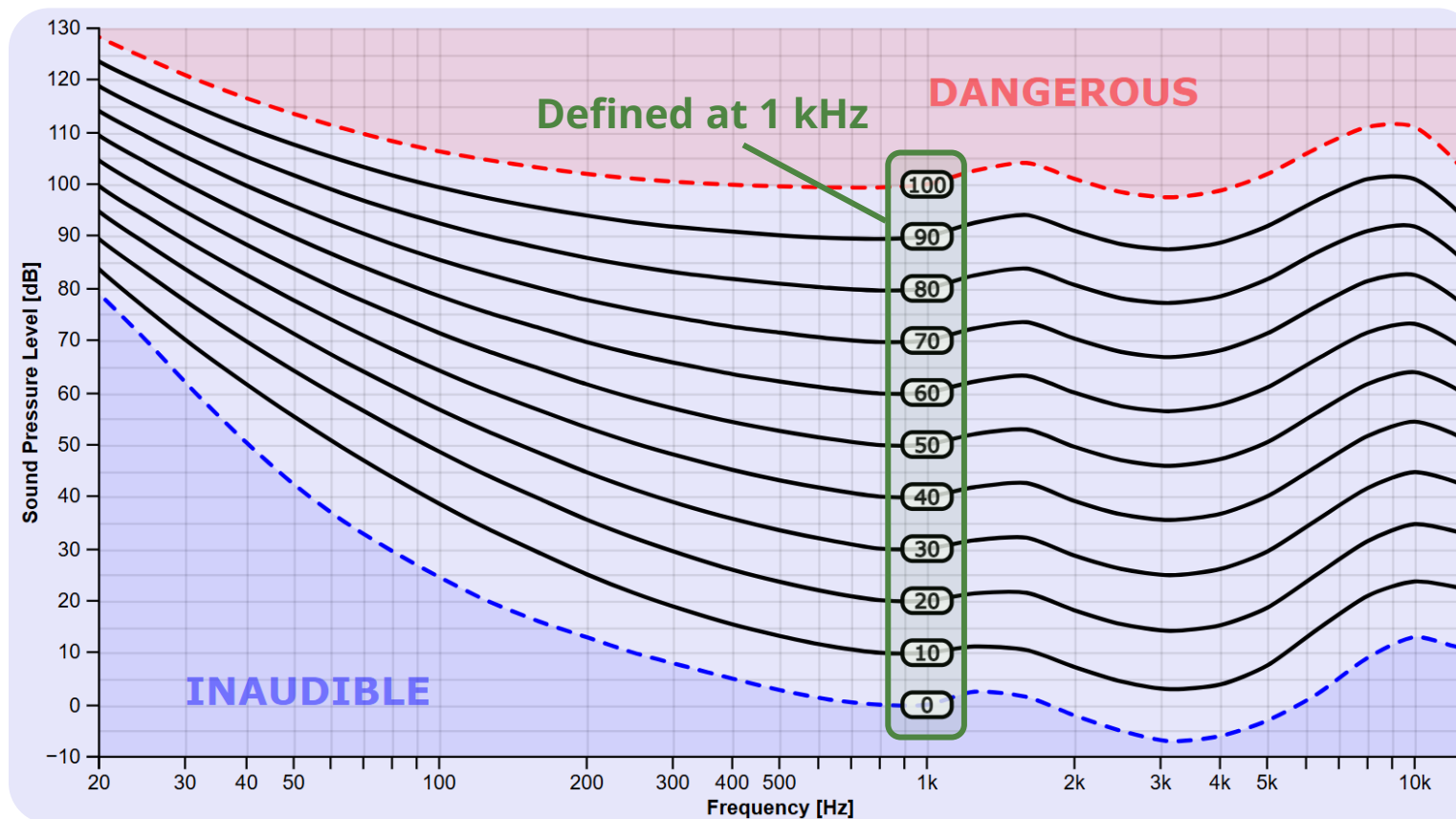
(Source: Parker, 2024)

# Loudness Perception: Equal-loudness Contours



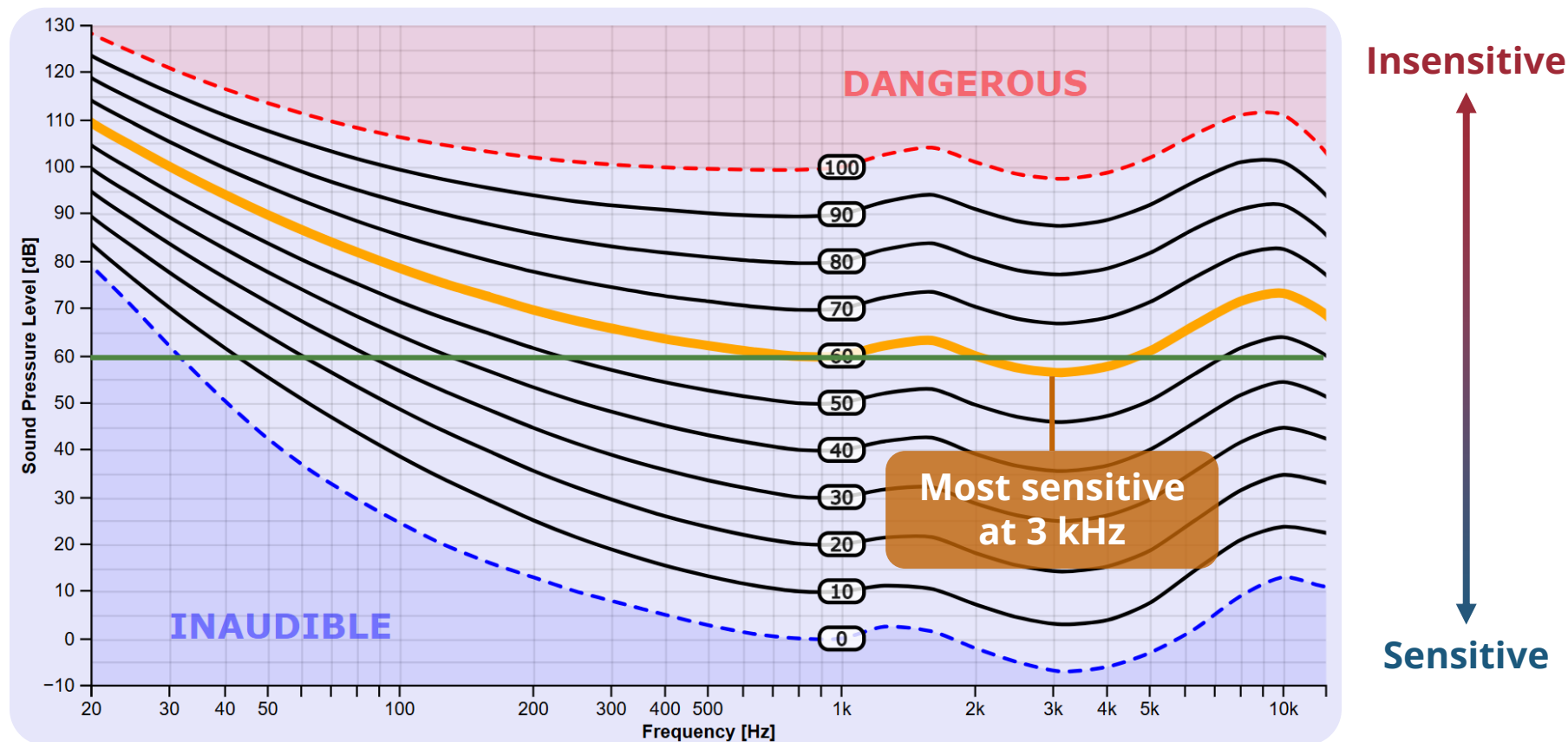
(Source: Parker, 2024)

# Loudness Perception: Equal-loudness Contours



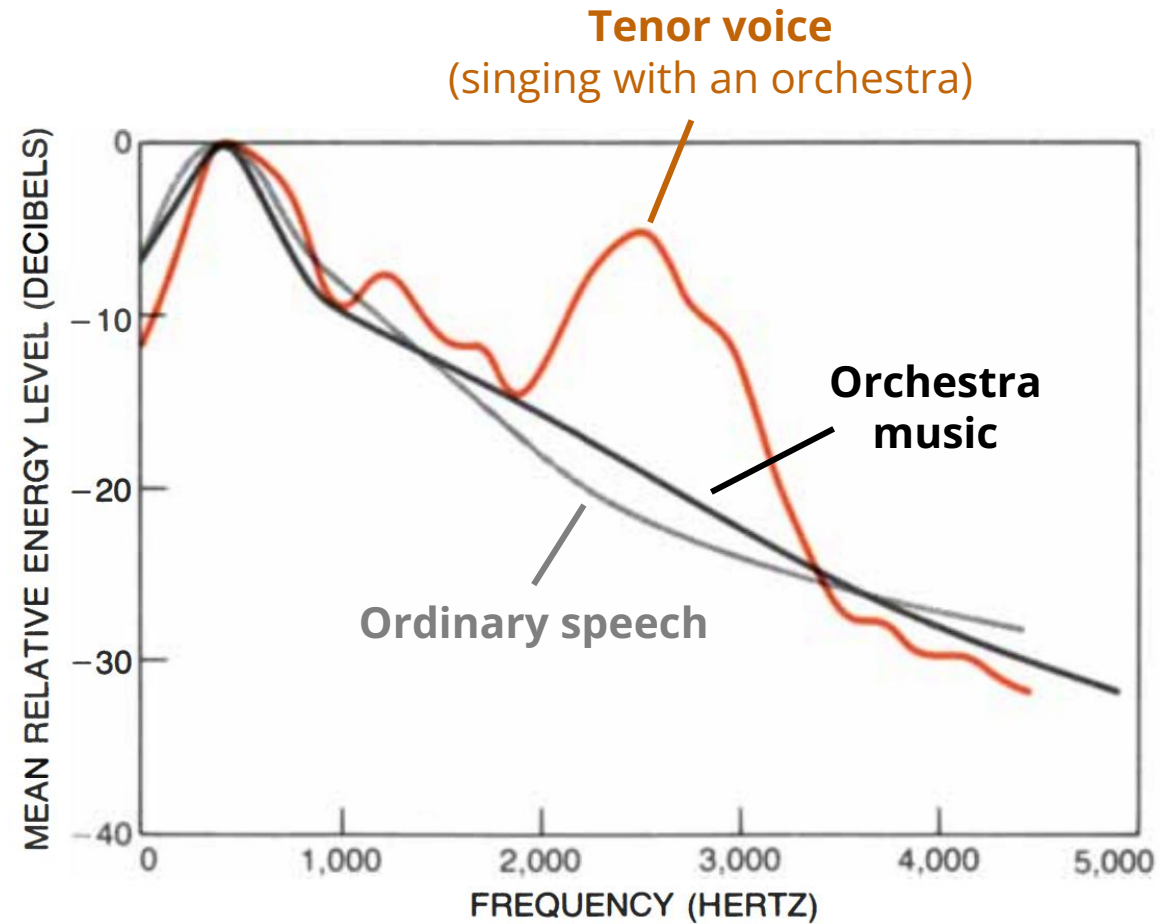
(Source: Parker, 2024)

# Loudness Perception: Equal-loudness Contours



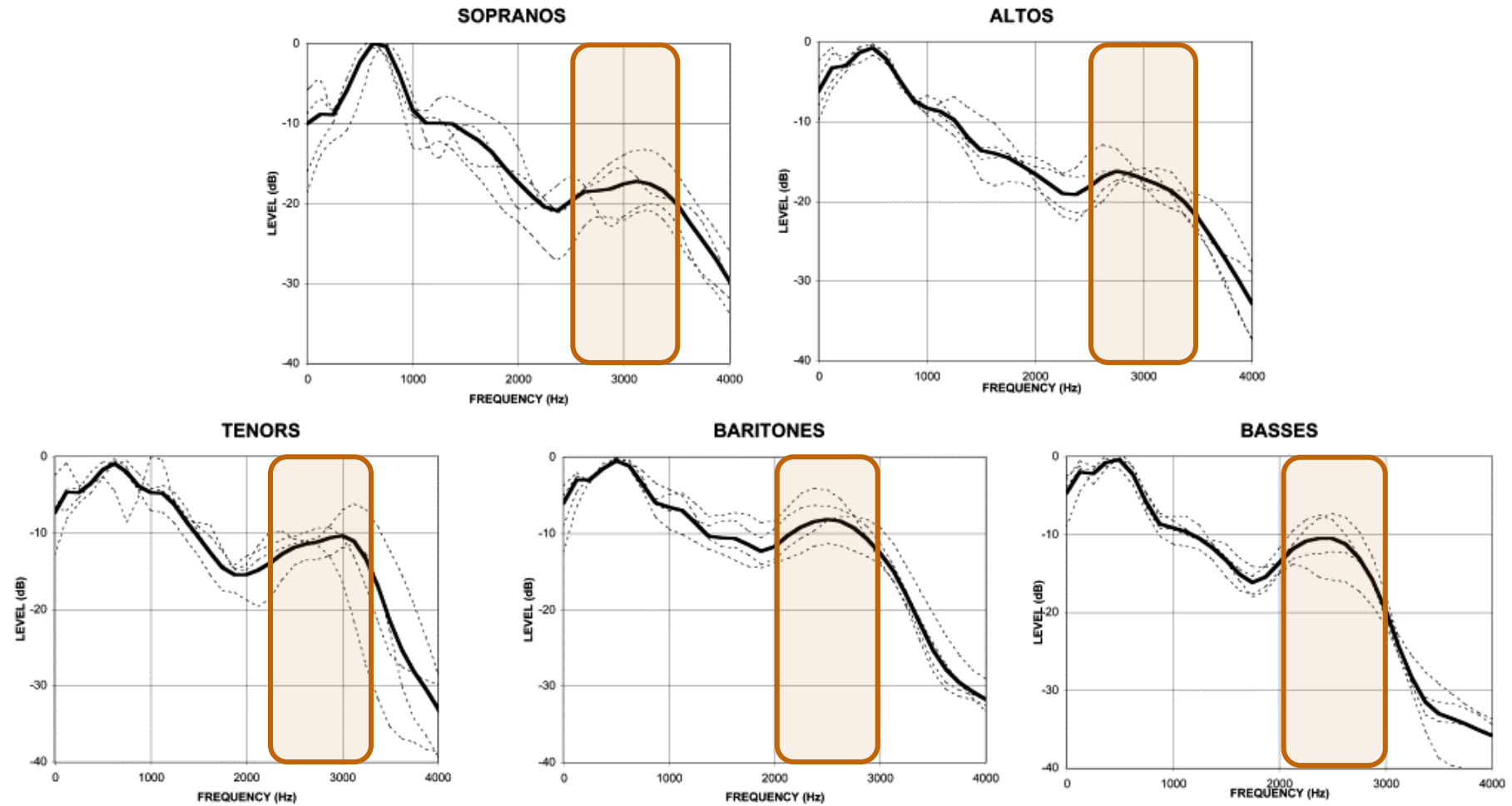
(Source: Parker, 2024)

# Singer's Formants (Sundberg, 1991)



(Source: Sundberg, 1977)

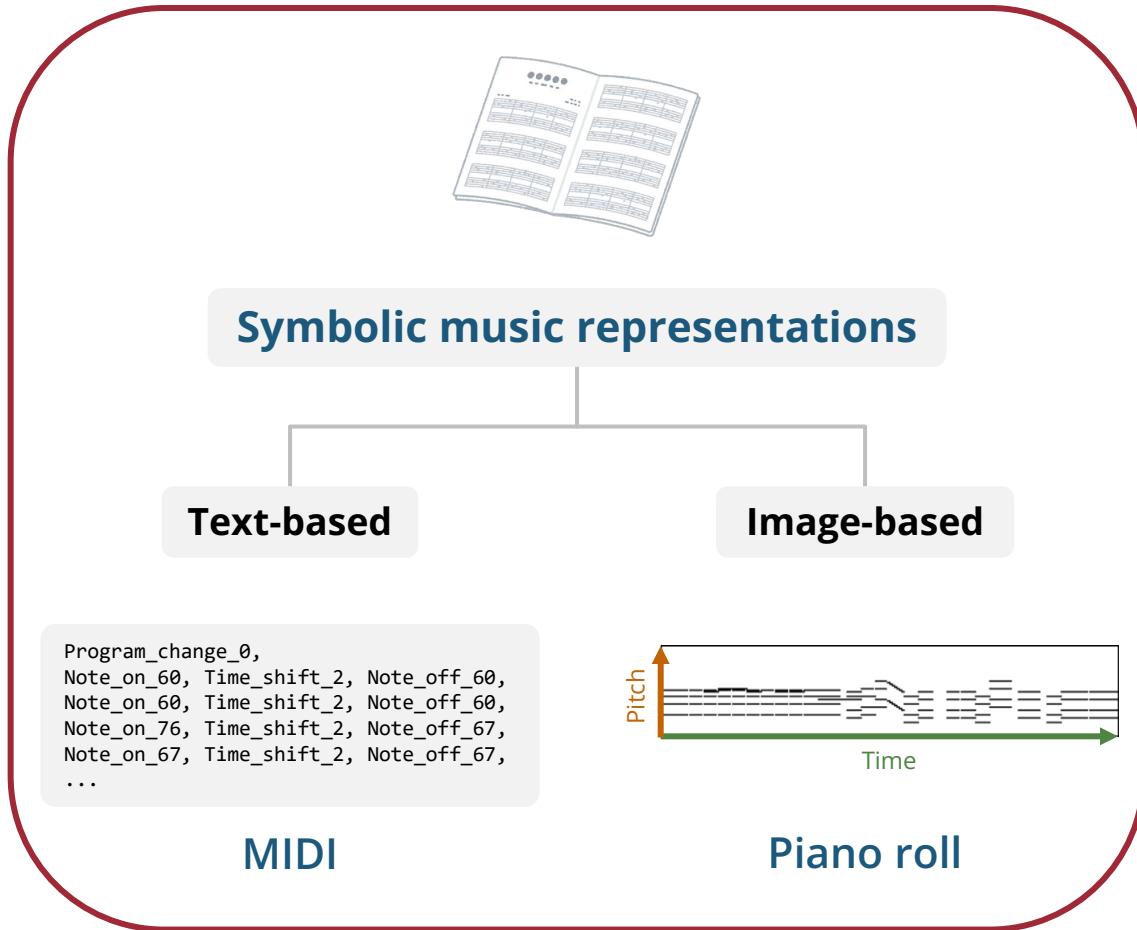
# Singer's Formants (Sundberg, 1991)



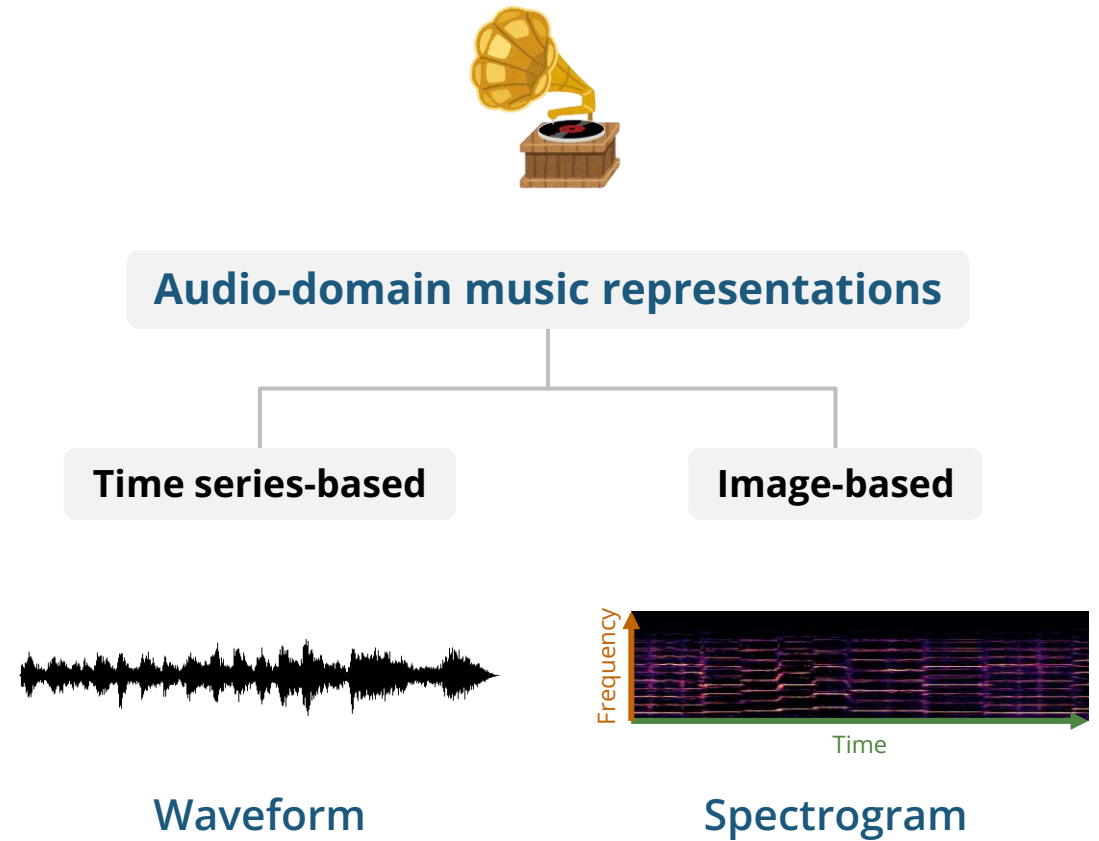
(Source: Sundberg, 1977)

# Recap

# Four Representative Music Representations



**Today's topic!**



# An Example of ABC Notation

Ah! vous dirai-je, maman  
(Twinkle, twinkle, little star)

*anon. (France)*

♩ = 120

Metadata

```
X:571
T:Ah! vous dirai-je, maman
T:(Twinkle, twinkle, little star)
C:anon.
O:France
R:Nursery song
M:C Meter
L:1/4 Unit note length (temporal resolution)
Q:120 Tempo
K:C Key
CCGG|AAG2|FFEE|DDC2:|
|:GGFF|EED2|GGFF|EED2|
CCGG|AAG2|FFEE|DDC2:|
```

# Representing Music using MIDI Messages

- Three main MIDI messages
  - Note on
  - Note off
  - Time Shift

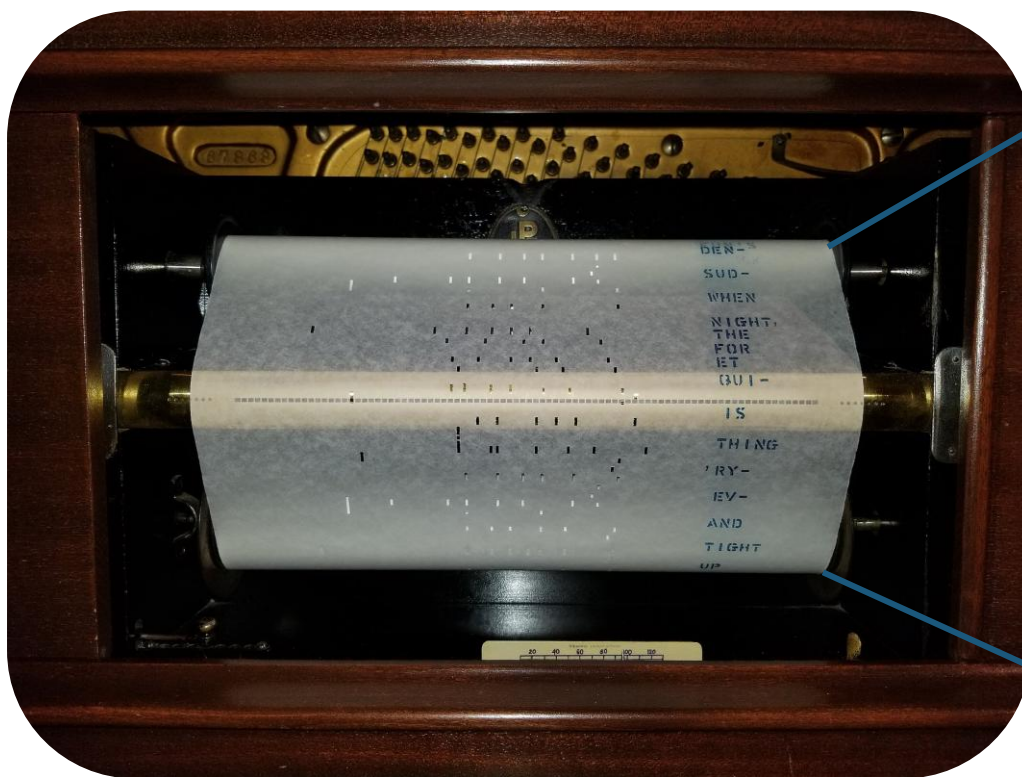
Sunshine on the Meadow



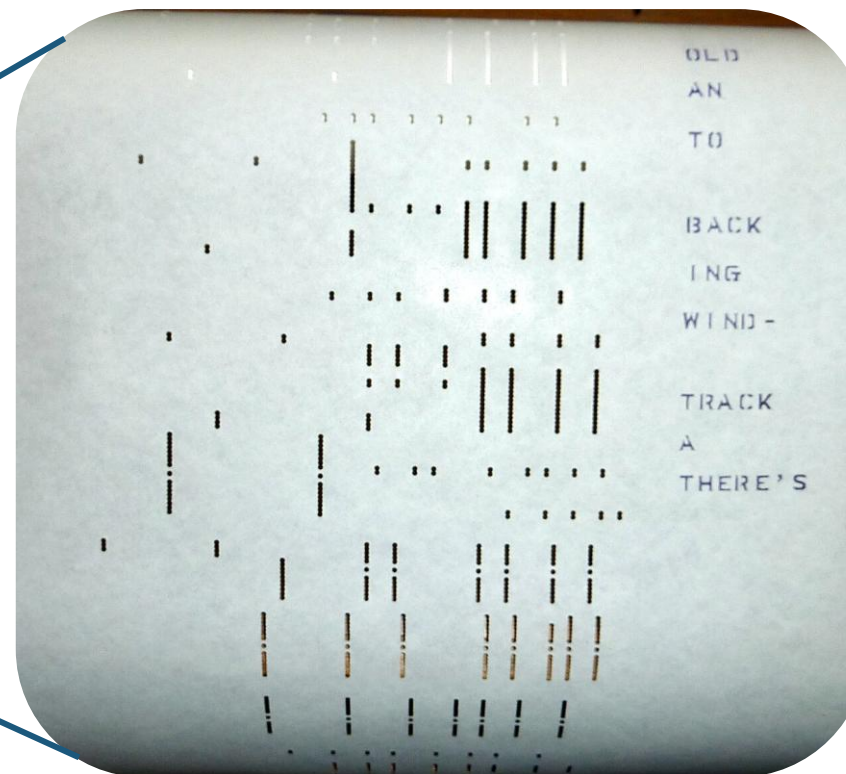
The image shows two staves of musical notation. The top staff is in 4/4 time and features a treble clef. The first two notes are circled: the first note (G4) is circled in blue with an orange 'X' over it, and the second note (A4) is circled in green. A red arrow points from the first note to the second, indicating a time shift. The rest of the staff contains a sequence of quarter notes: B4, C5, D5, E5, F5, G5, A5, B5, C6, D6, E6, F6, G6, A6, B6, C7. The bottom staff continues the melody with quarter notes: D6, E6, F6, G6, A6, B6, C7, D7, E7, F7, G7, A7, B7, C8, D8, E8, F8, G8, A8, B8, C9.

```
Note_on_67 Time_shift_quarter_note Note_off_67  
Note_on_67 Time_shift_quarter_note, Note_off_67,  
Note_on_64, Time_shift_quarter_note, Note_off_64,  
Note_on_64, Time_shift_quarter_note, Note_off_64,  
...
```

# Piano Rolls

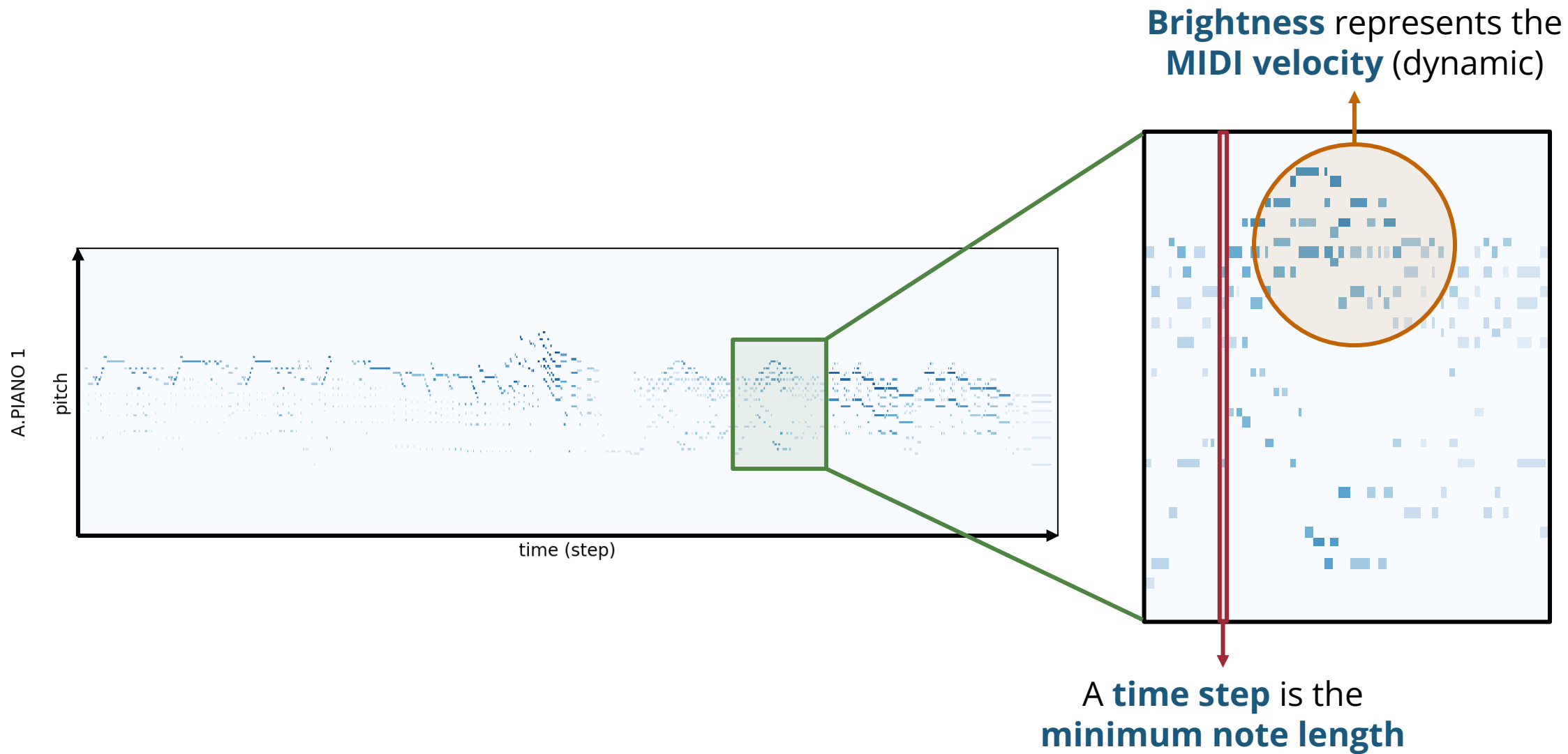


(Source: Draconichiaro)

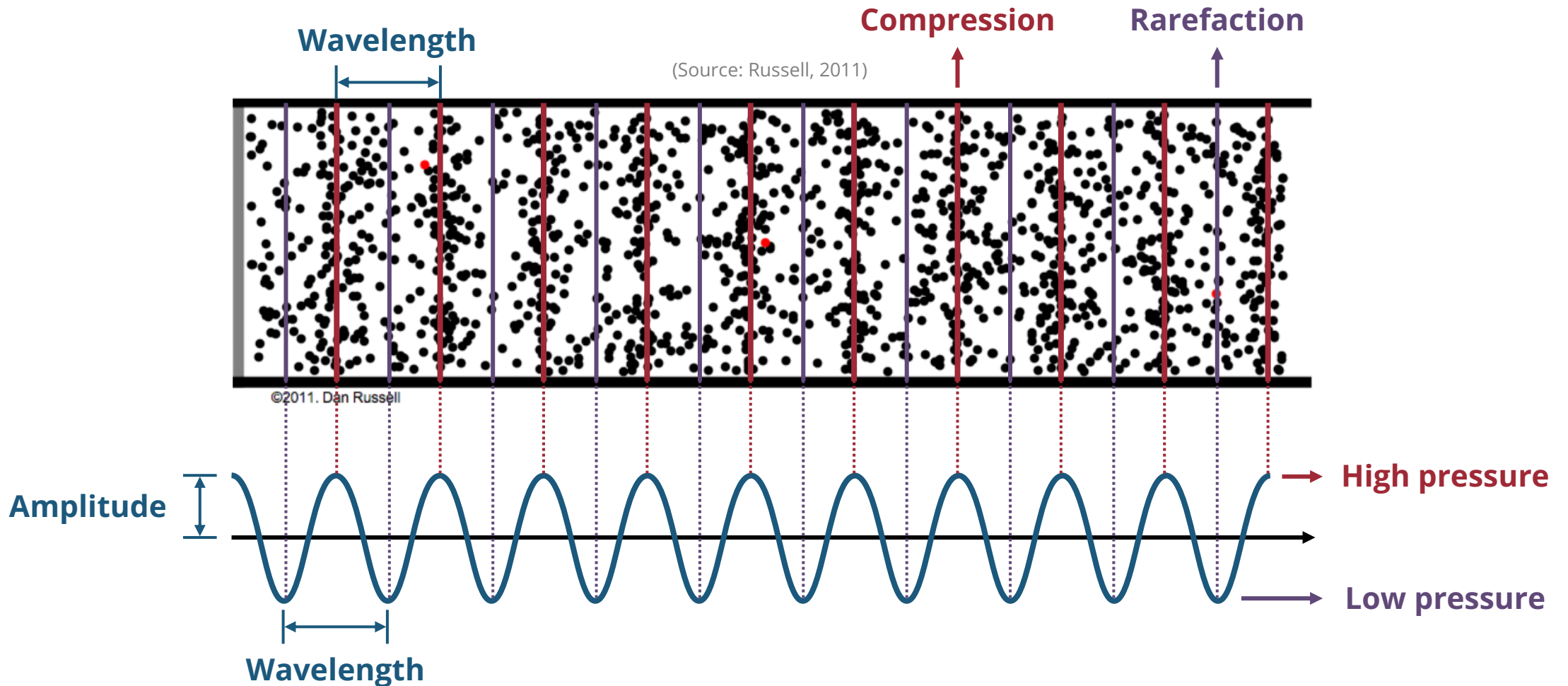


(Source: Tangerineduel)

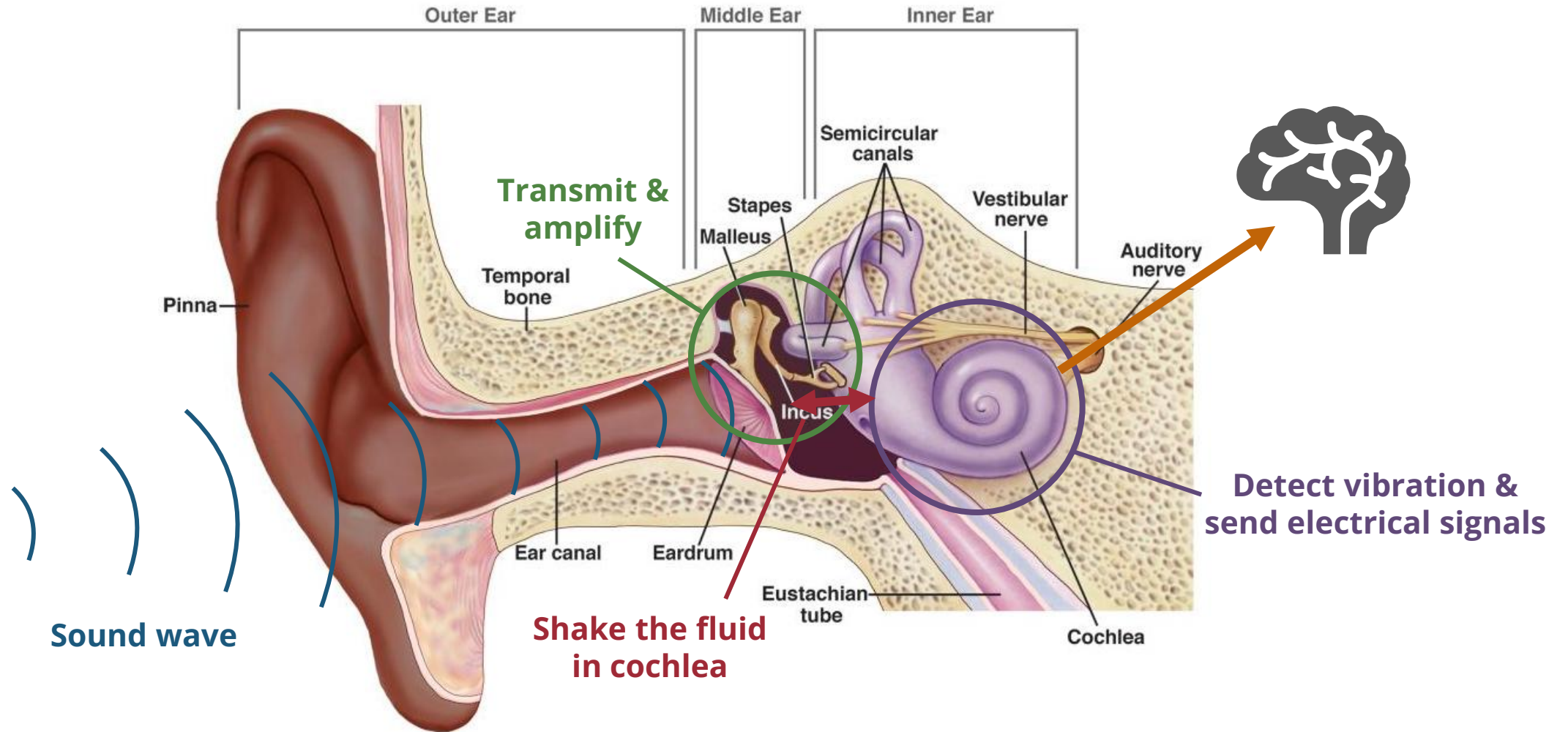
# Piano Roll Representation



# Longitudinal vs Transverse Waves

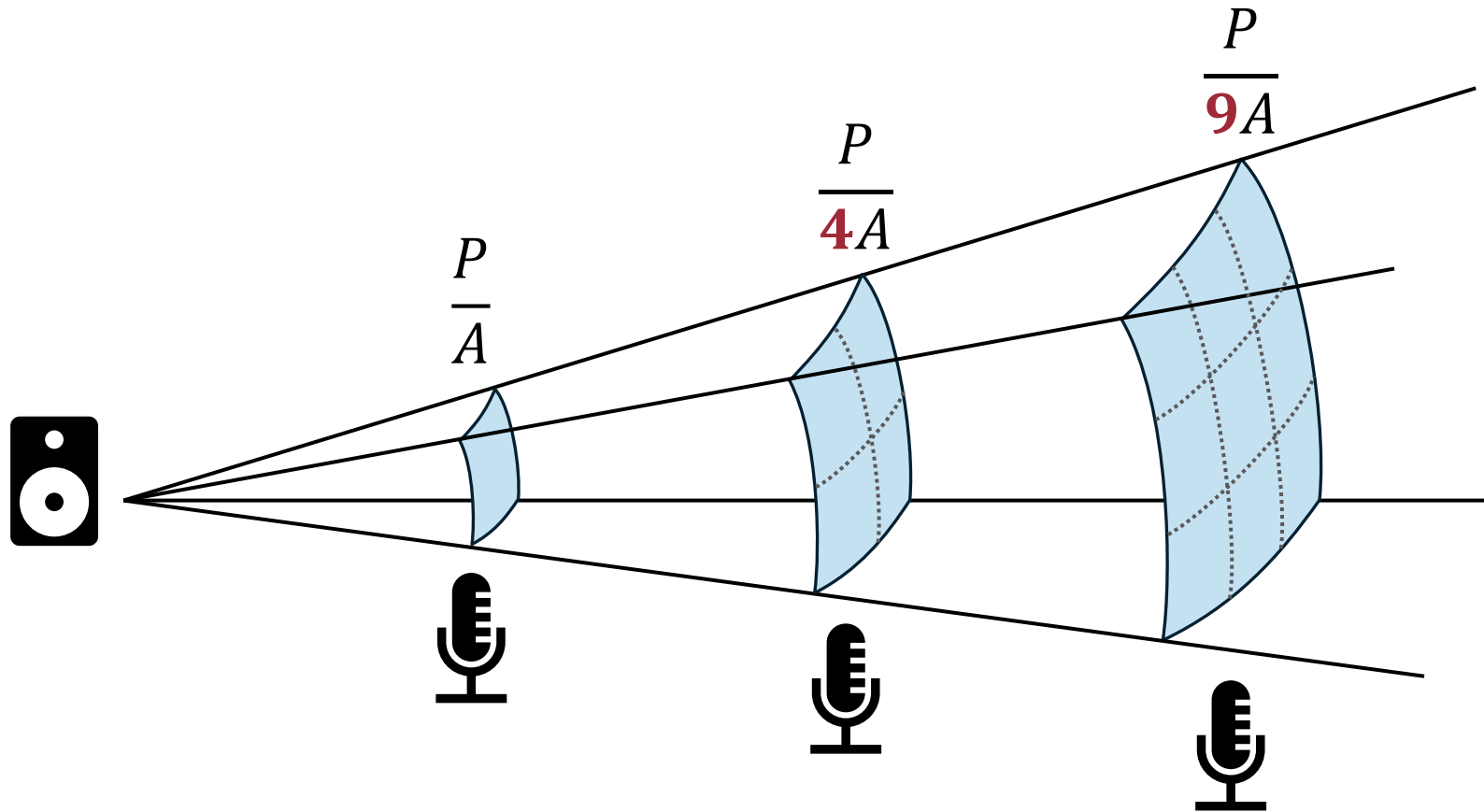


# Human Ears



(Source: NIH/NIDCD)

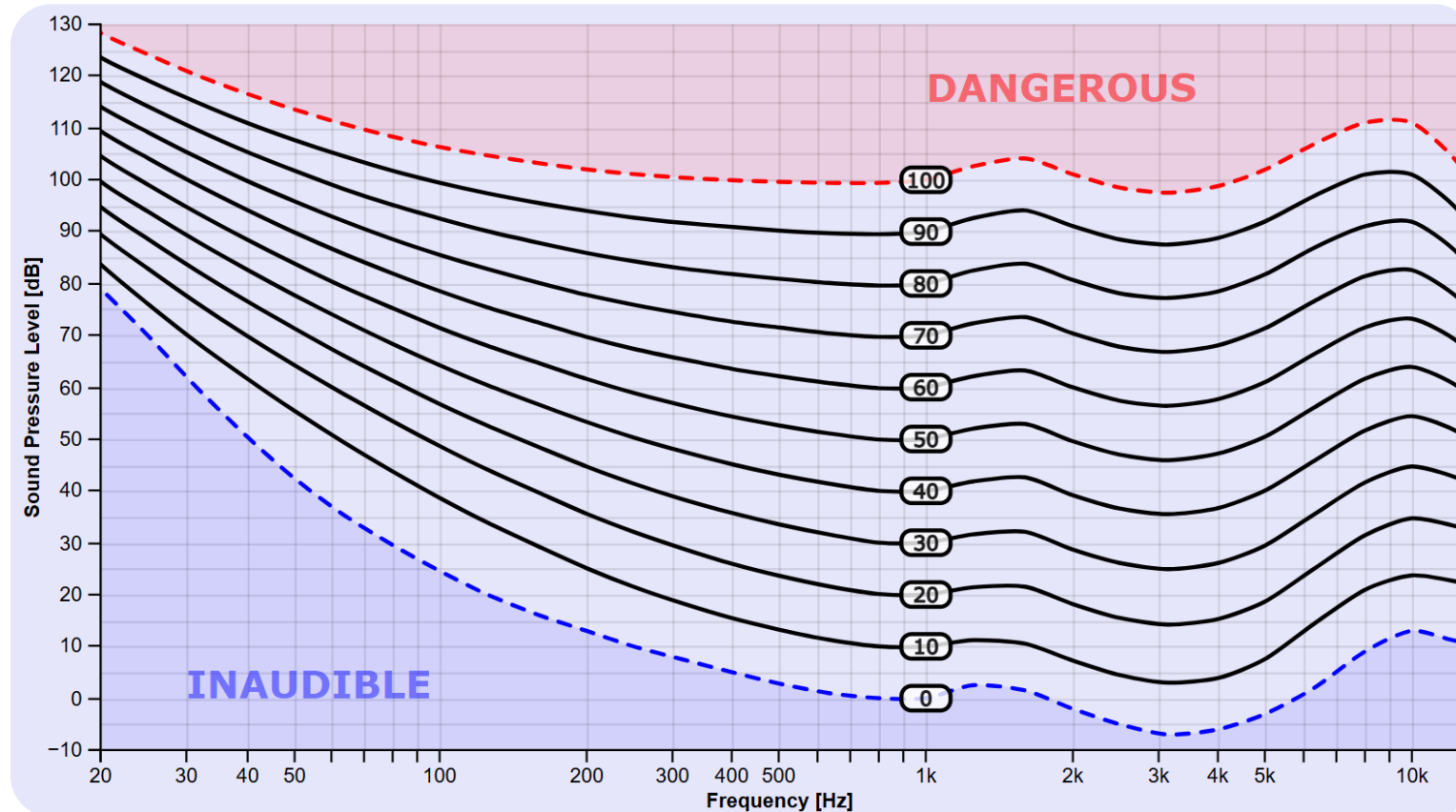
# Sound Propagation & Inverse Square Law



Inverse Square Law

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

# Loudness Perception: Equal-loudness Contours

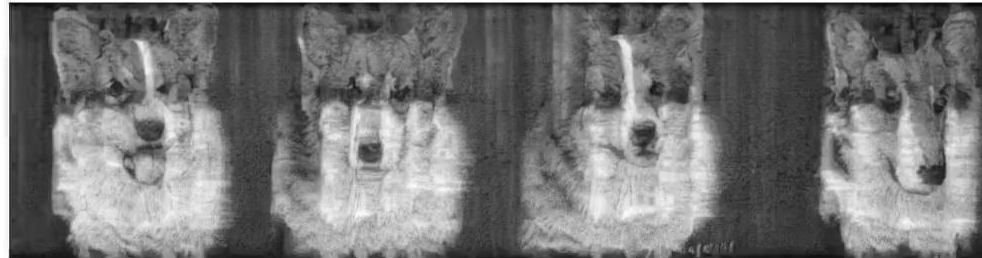
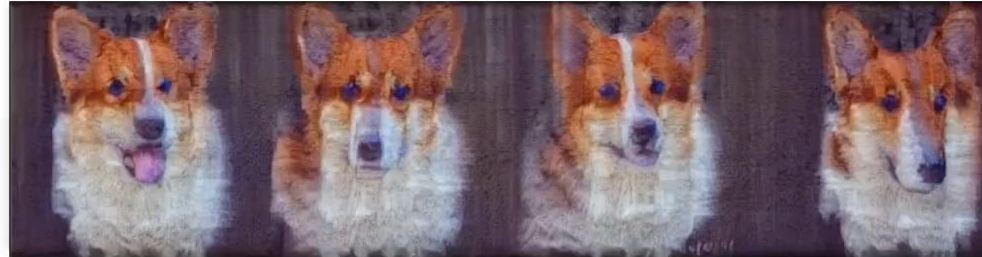


(Source: Parker, 2024)

## Next Lecture

# Audio Processing Fundamentals

Image prompt: a colorful photo of corgis



Audio prompt: dog barking

(Source: Chen et al., 2024)