

PAT 463/563 (Fall 2025)

Music & AI

Lecture 4: Music Processing Fundamentals

Instructor: Hao-Wen Dong

Four Representative Music Representations



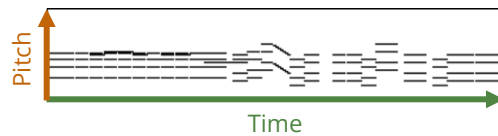
Symbolic music representations

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



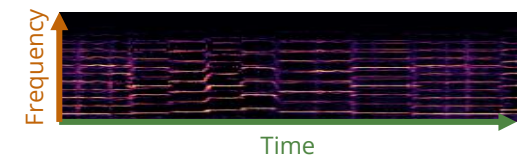
Audio-domain music representations

Time series-based



Waveform

Image-based



Spectrogram

Today's topic!

Machine-readable Music Language

Designing a Machine-readable Music Language

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



- A simple text-based notation
- Use **letters** to denote **pitches**
 - 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 : |
```

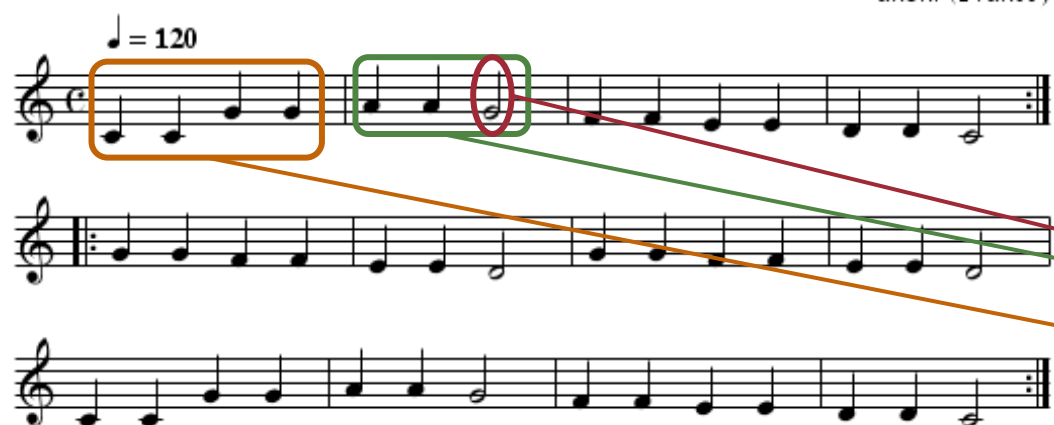
Twinkle, twinkle, little star!

An Example of ABC Notation

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

anon. (France)

$\text{♩} = 120$



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!

SA 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

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



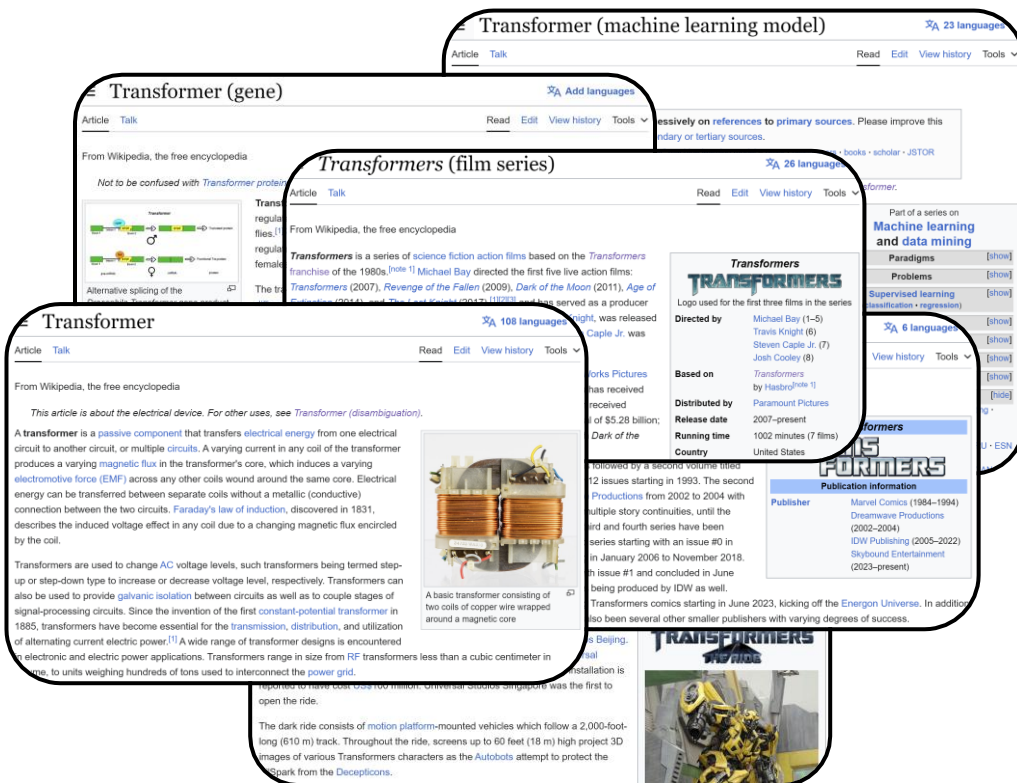
ChatGPT

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



A transformer is a _____

electrical device ↑

fiction character ↑

deep learning model ↑

family of genes ↑

type of food ↓

musical instrument ↓

| Language Models (Mathematically)

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

$$P(\underset{\substack{\text{Next word} \\ \swarrow}}{x_i} \mid \underbrace{x_1, x_2, \dots, x_{i-1}}_{\substack{\text{Previous words} \\ \searrow}})$$

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

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

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

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

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

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

Music Language Models (Mathematically)

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

$$P(\underset{\substack{\text{Next note}}}{x_i} \mid \underbrace{x_1, x_2, \dots, x_{i-1}}_{\substack{\text{Previous notes}}})$$

$$\begin{array}{l} P(\text{G} \mid C C G G A A) \uparrow \\ P(\text{A} \mid C C G G A A) \uparrow \\ P(\text{C} \mid C C G G A A) \uparrow \\ P(\text{F} \mid C C G G A A) \uparrow \\ P(\text{A}\flat \mid C C G G A A) \downarrow \\ P(\text{A}\sharp \mid C C G G A A) \downarrow \end{array}$$

| Language Models: Generation

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

A transformer is a

→ Model → deep

A transformer is a deep

→ Model → learning

A transformer is a deep learning

→ Model → model

A transformer is a deep learning model

→ Model → introduced

A transformer is a deep learning model introduced

→ Model → in

A transformer is a deep learning model introduced in

→ Model → 2017

| Limitations of ABC Notations

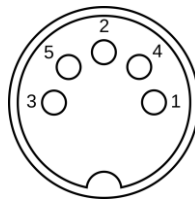
- Limited expressiveness
- Monophonic tunes only

MIDI-like Representation

MIDI (Musical Instrument Digital Interface)



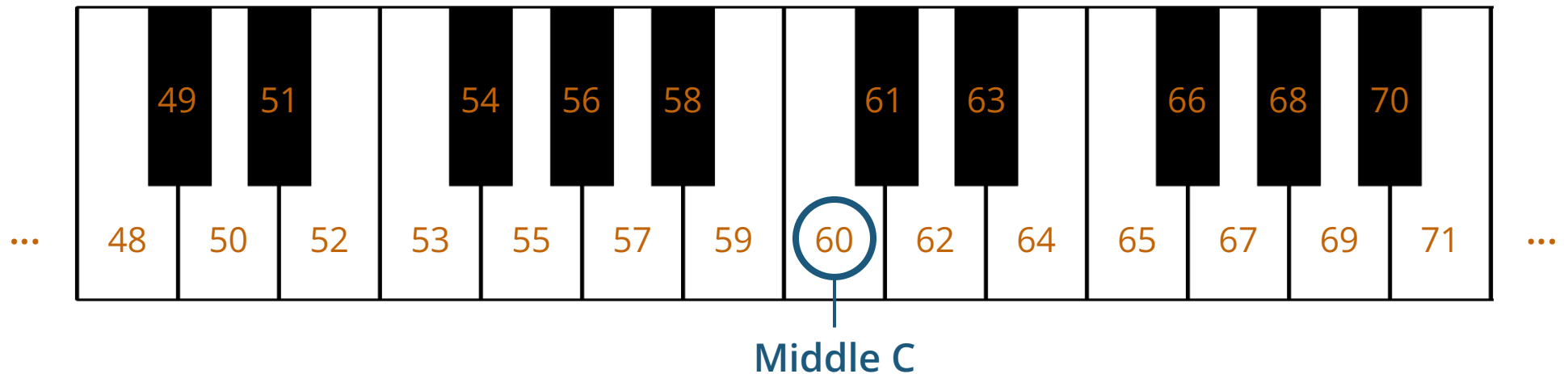
- A communication **protocol** between devices
- MIDI Messages
 - Note on
 - Note off
 - Delta time
 - Program change
 - Control change
 - Pitch bend change



MIDI I/O

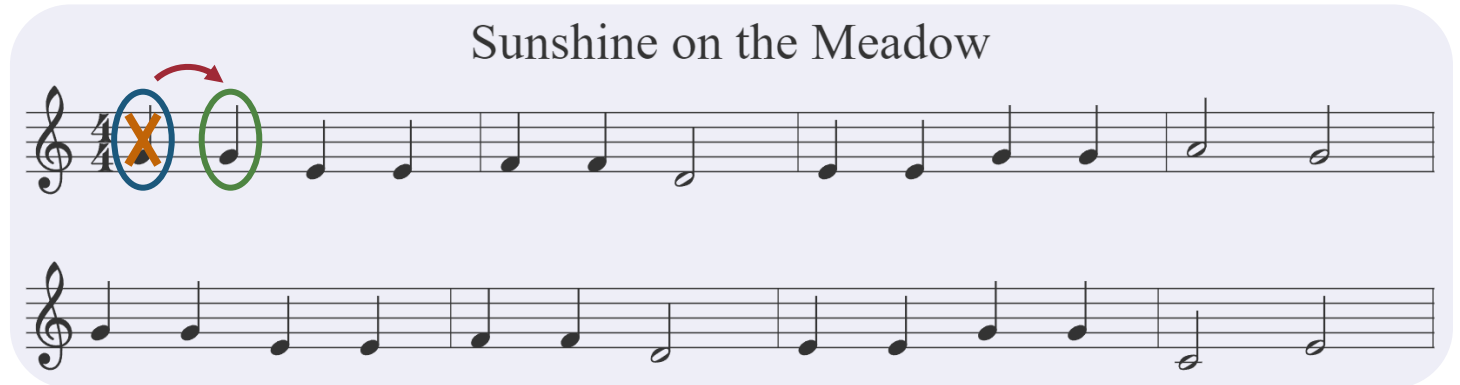
MIDI Note Numbers

- 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



```
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
3rd 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, ...

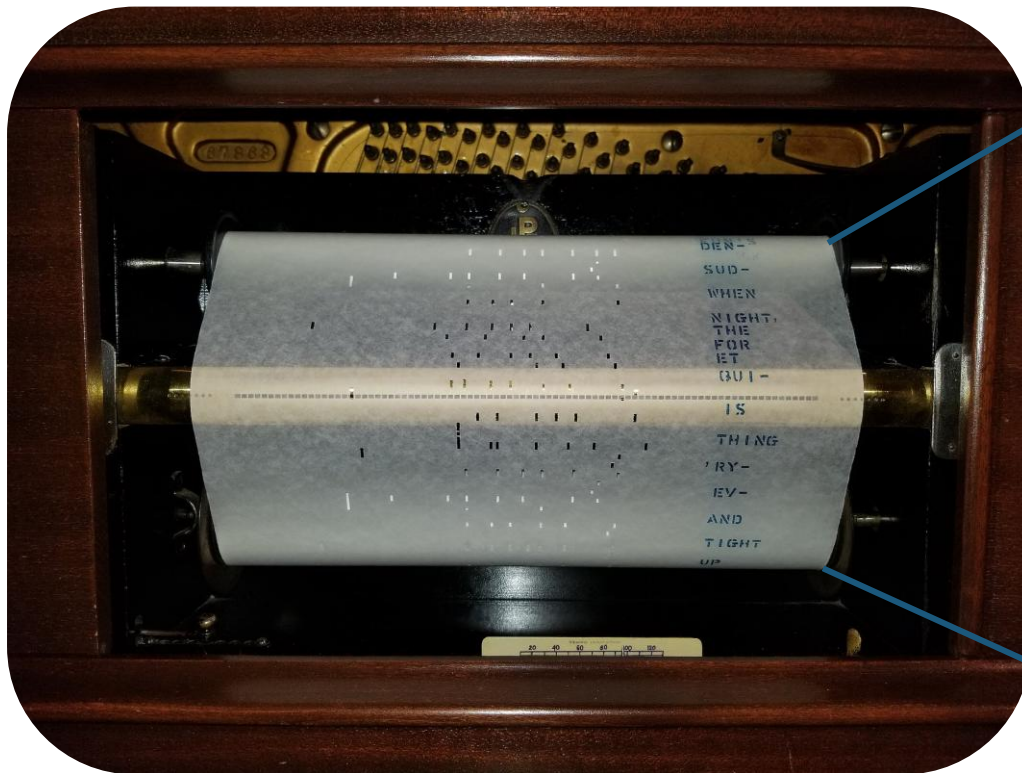
| A.I. Duet (Mann et al, 2016)



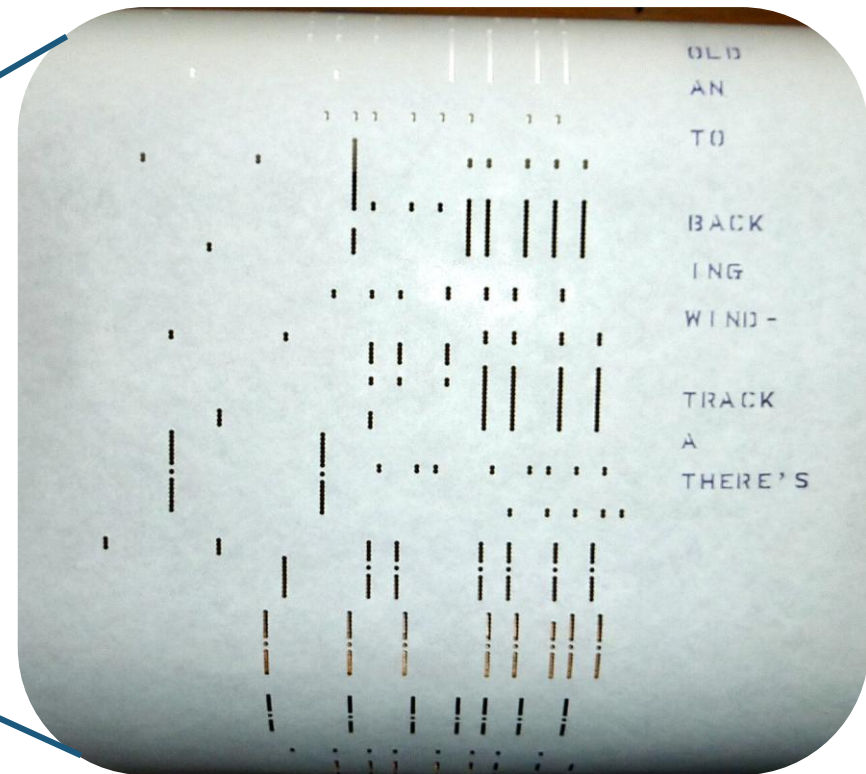
youtu.be/0ZE1bfPtvZo
experiments.withgoogle.com/ai/ai-duet/view

Piano Roll Representation

Piano Rolls



(Source: Draconichiaro)



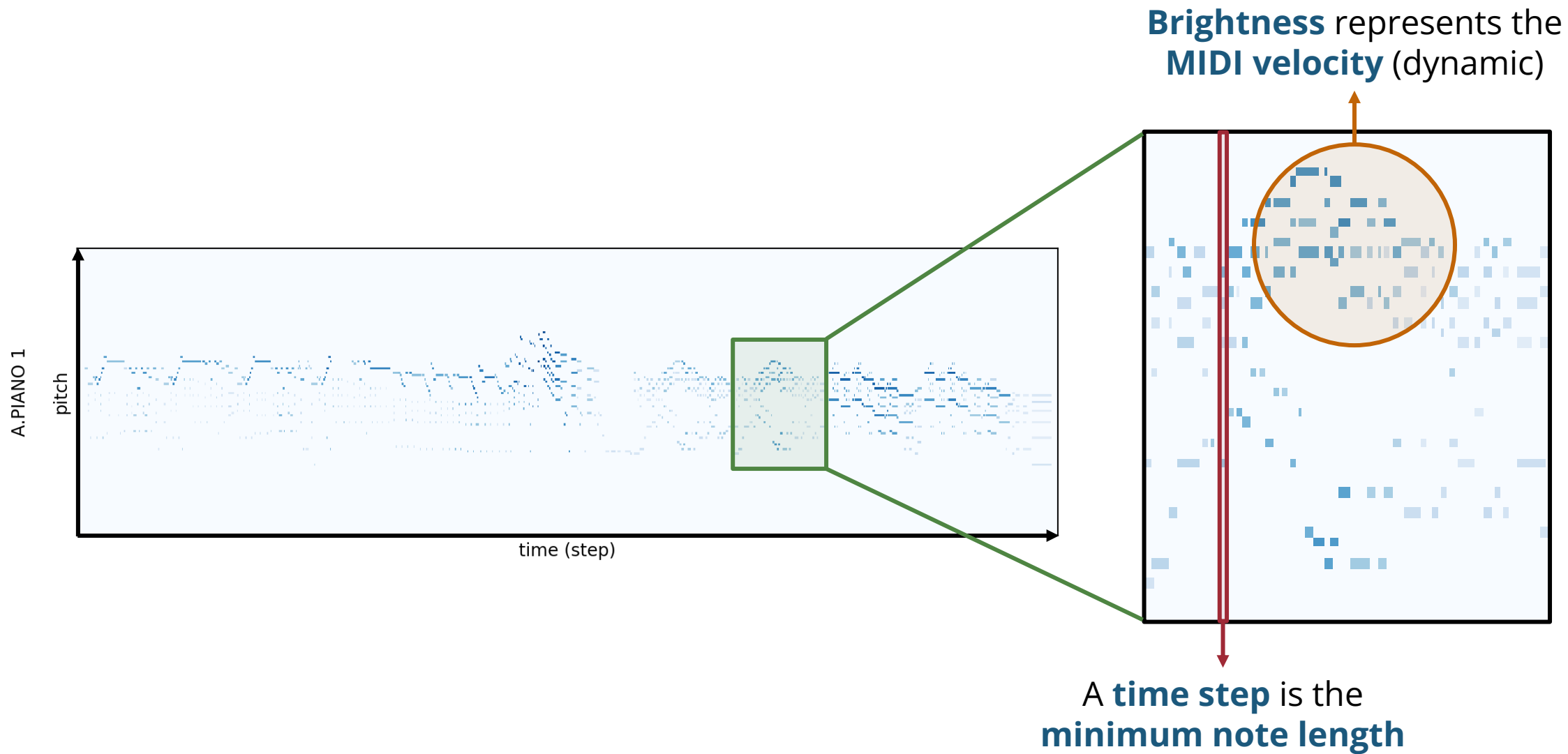
(Source: Tangerineduel)

| Player Pianos



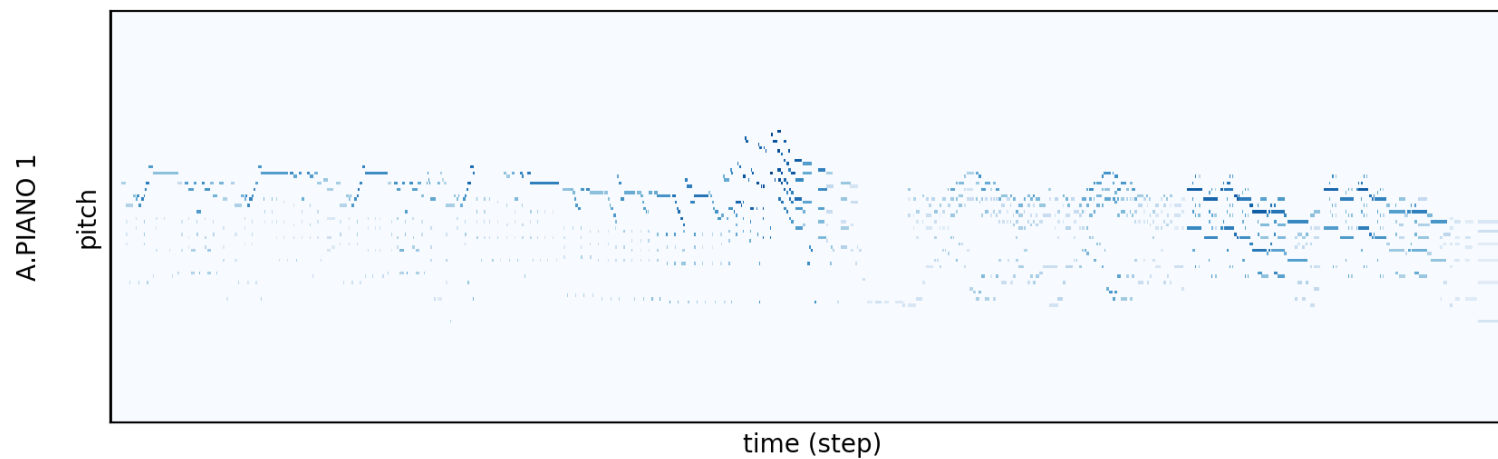
youtu.be/07krQ661fok

Piano Roll Representation

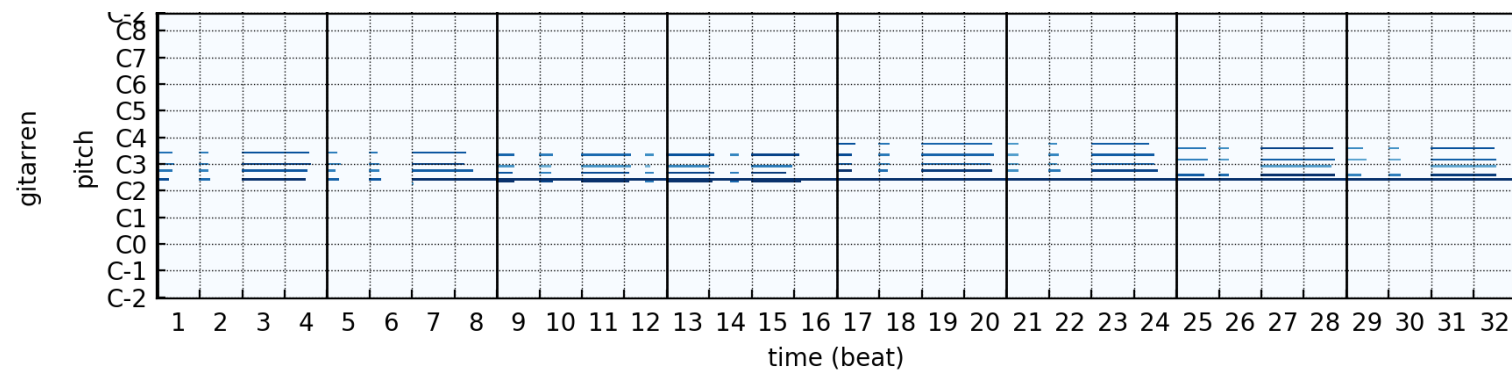


Piano Roll Representation

With expressive timing

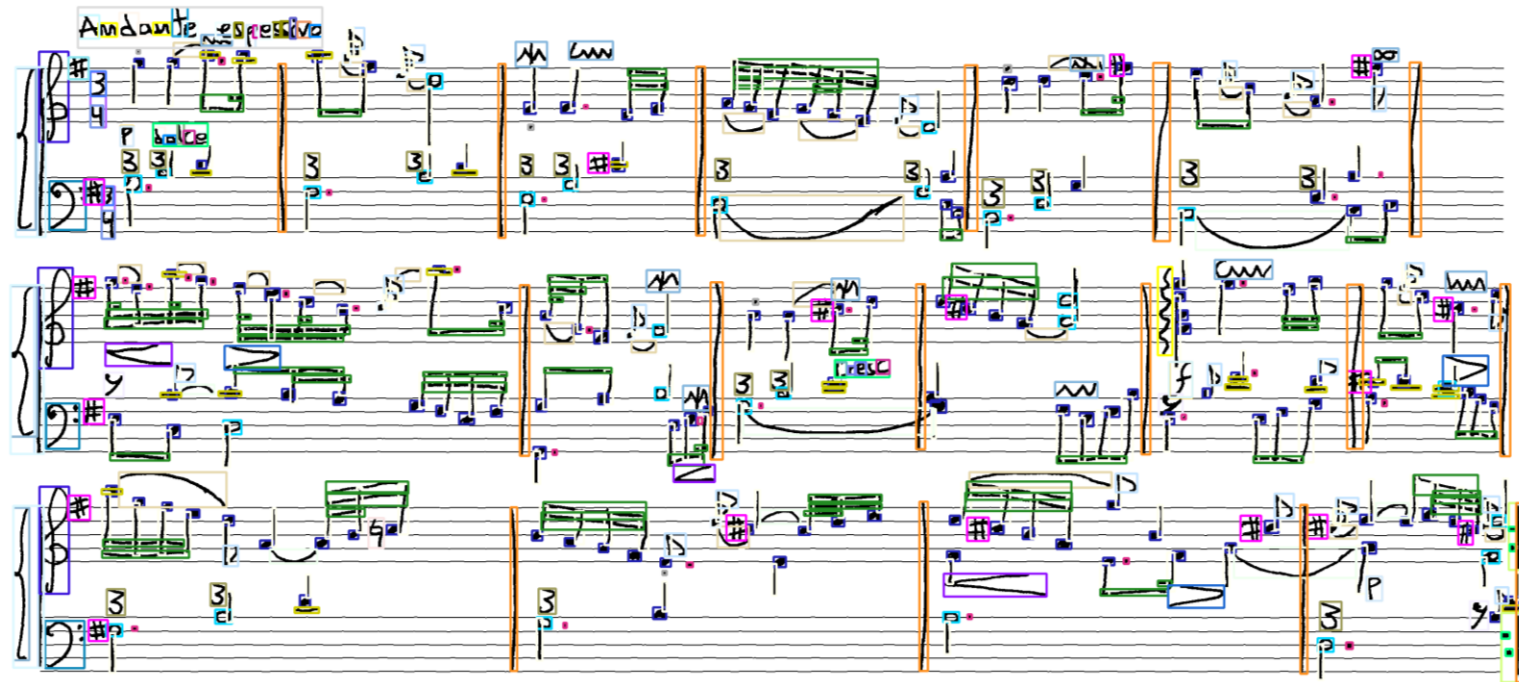


Without expressive timing



🤔 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)**

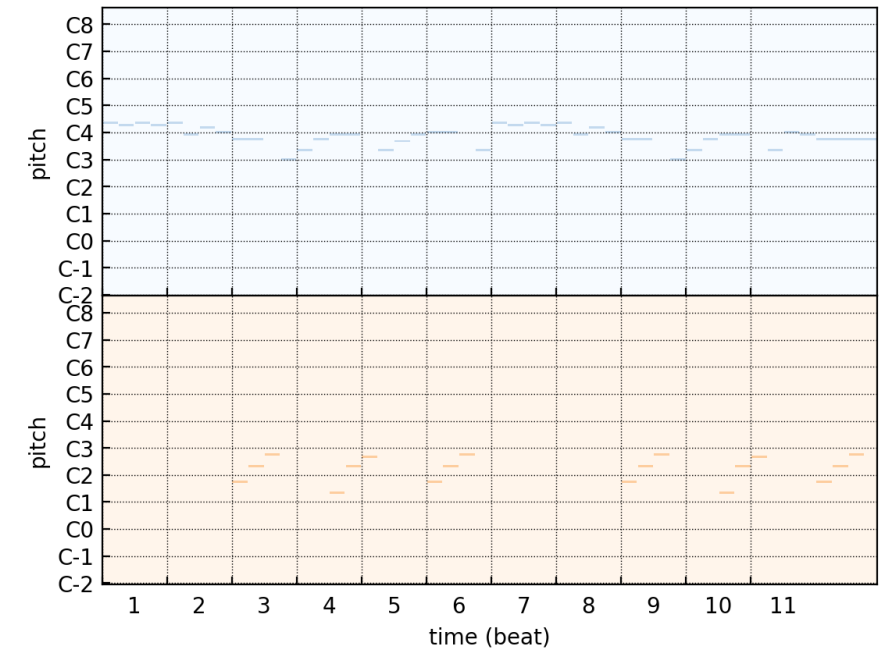


(Source: Pacha et al.)



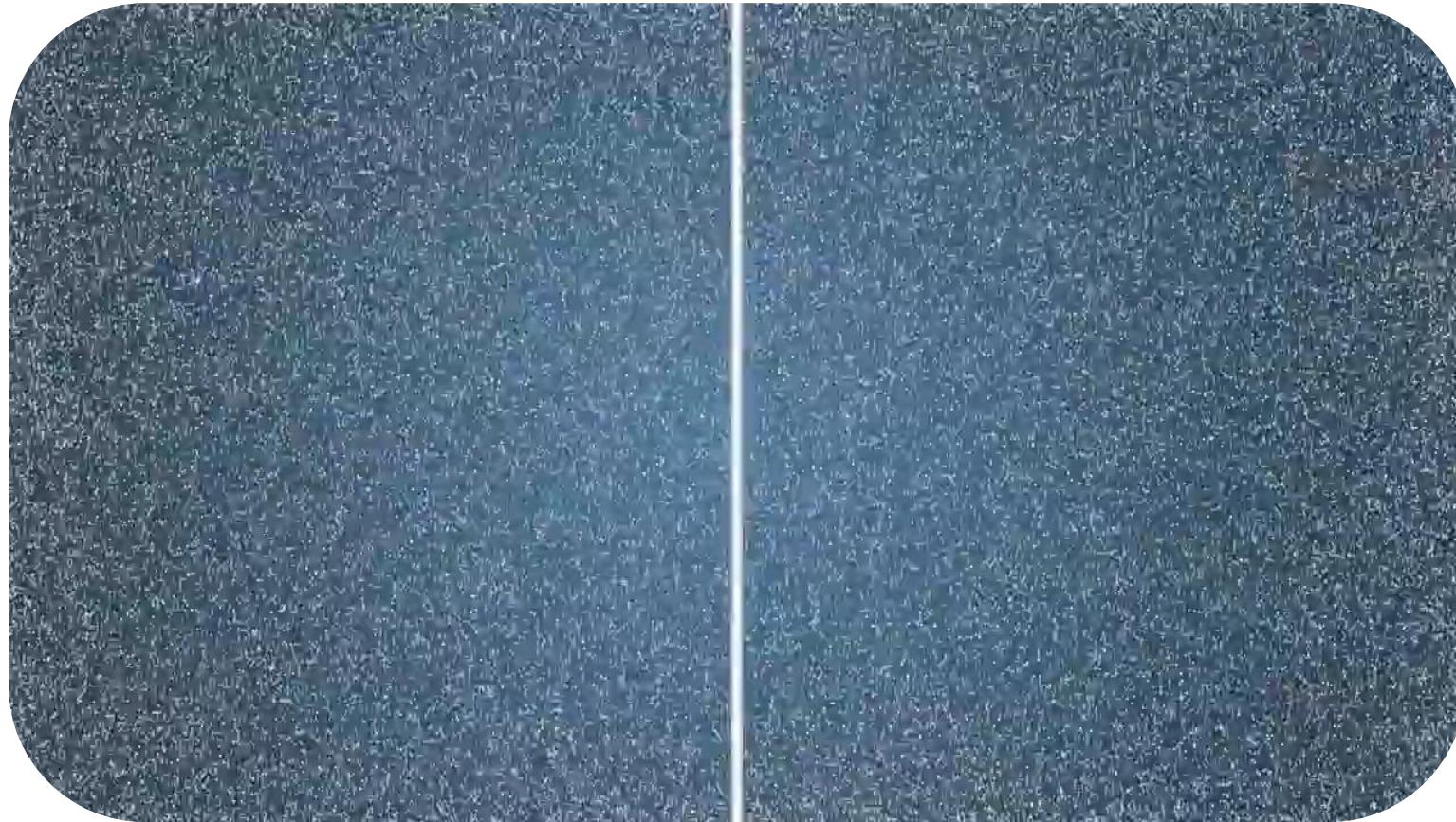
PA1: Symbolic Music Processing

- Instructions will be sent by **emails** and released on the **course website**
- 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**



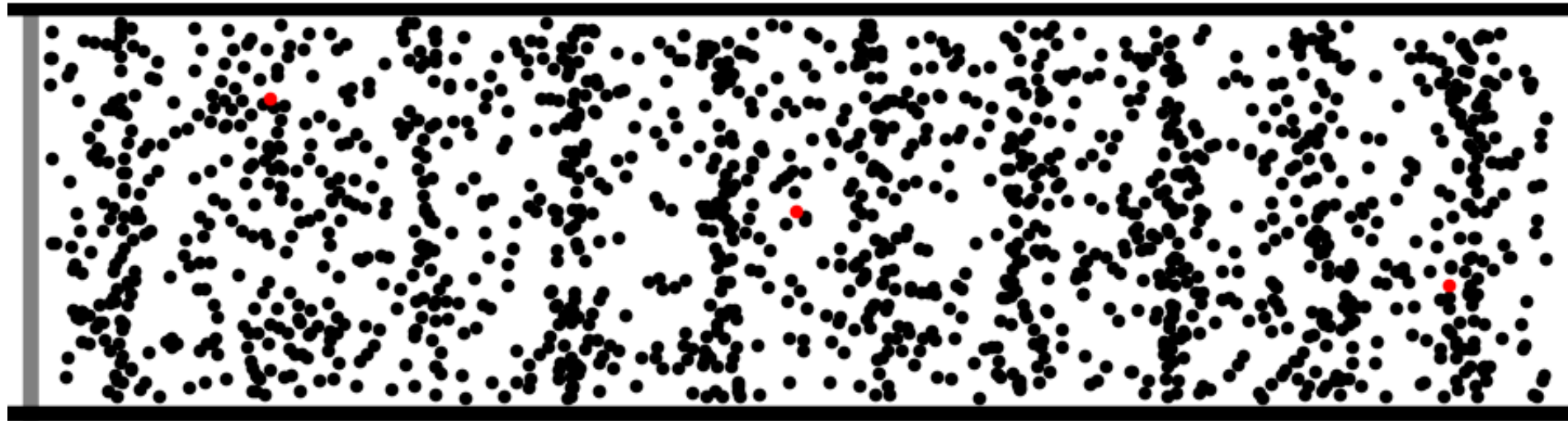
What is Sound?

| What is Sound?



youtu.be/aPswNDcteS4

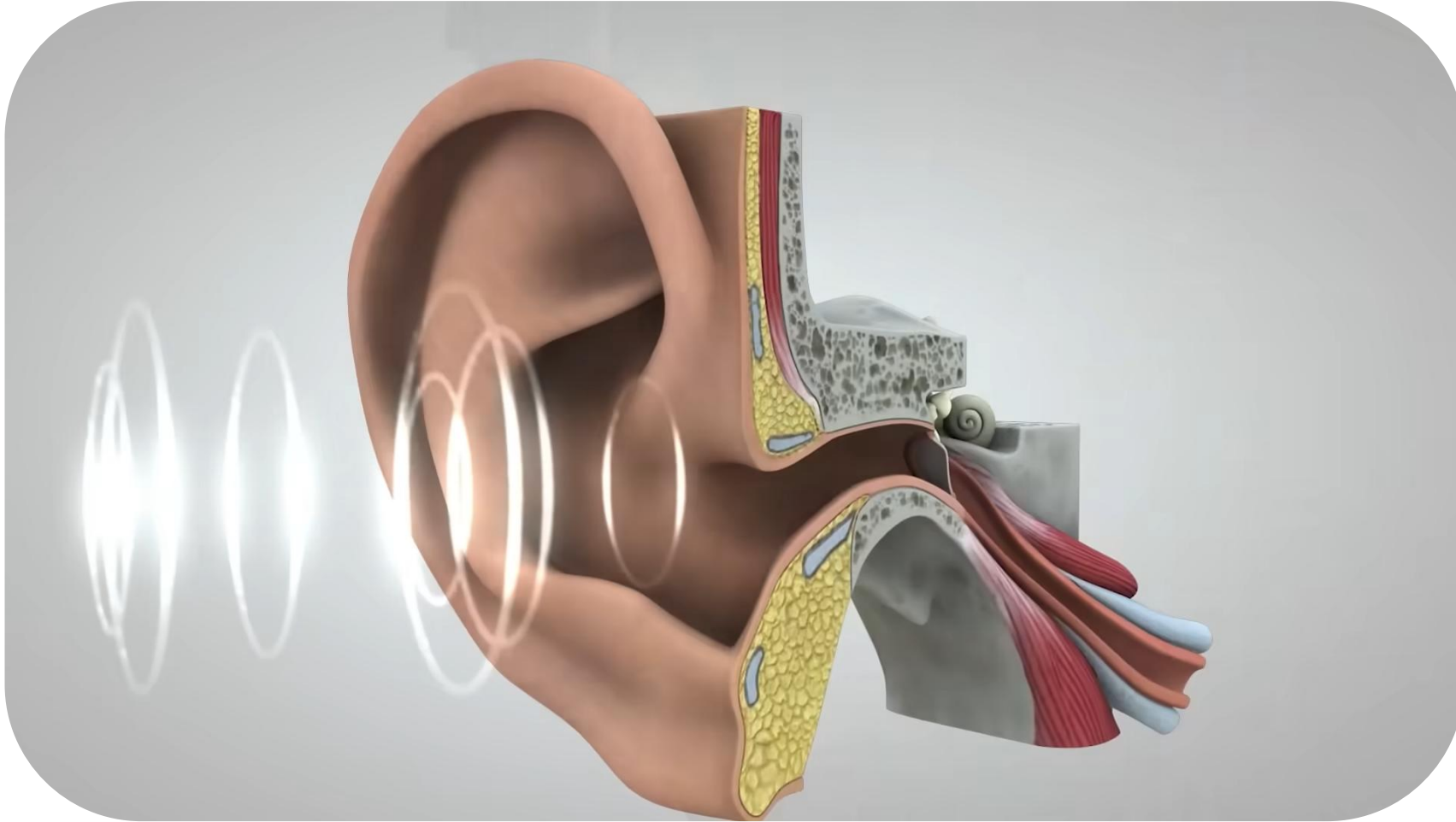
Sound is a Pressure Wave



©2011. Dan Russell

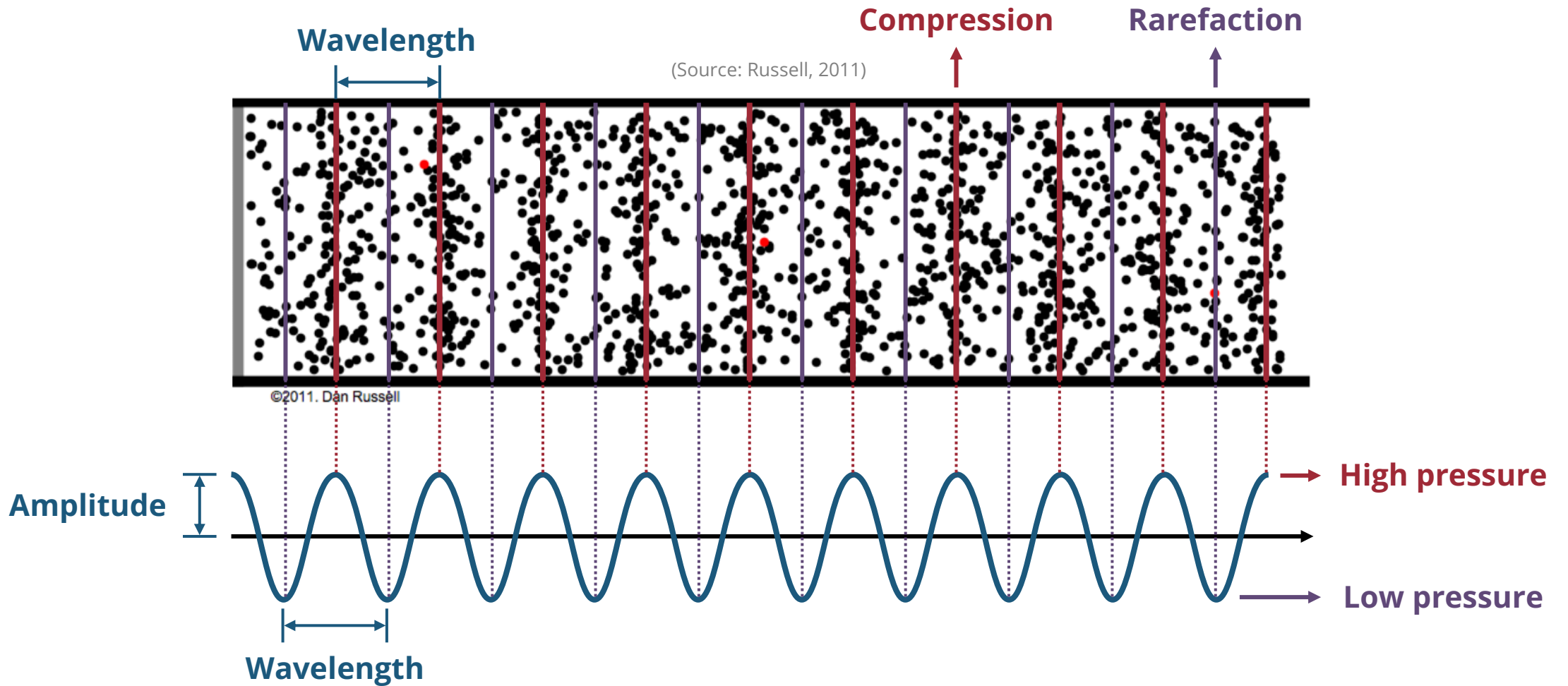
(Source: Russell, 2011)

| Human Auditory System



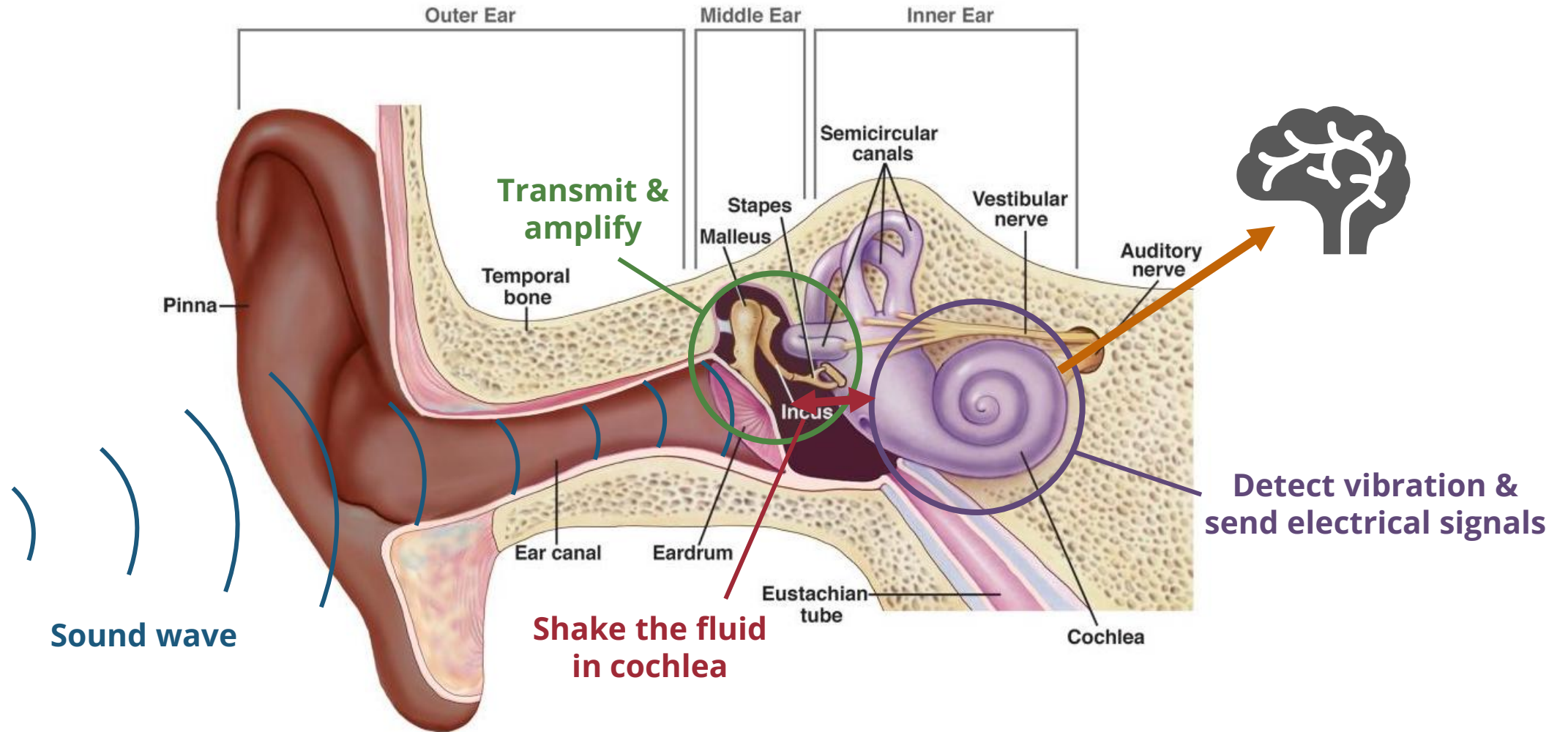
youtu.be/eQEaiZ2j9oc

Longitudinal vs Transverse Waves



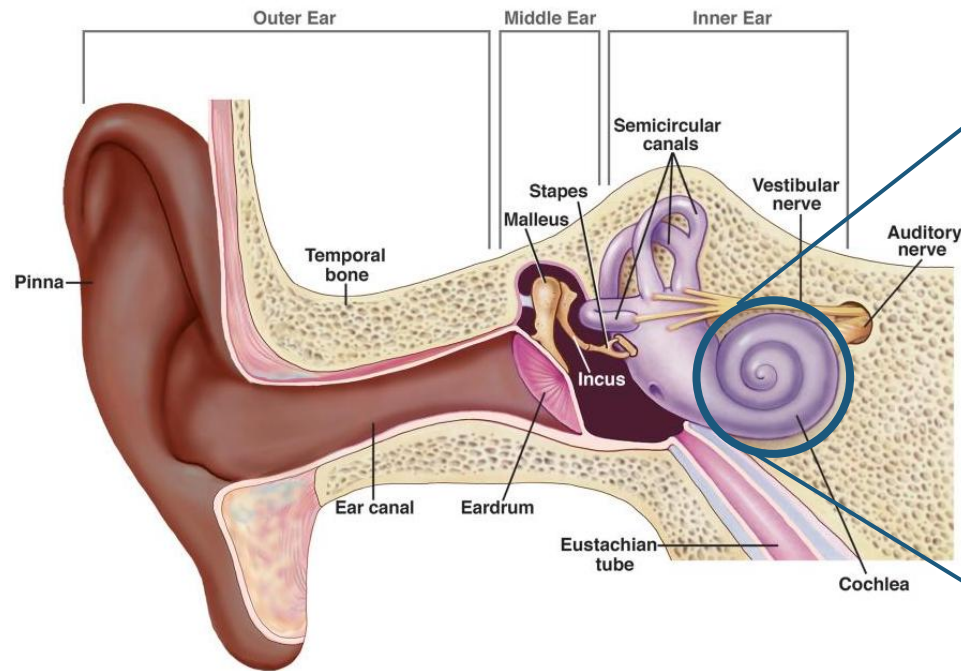
Auditory Perception

Human Ears

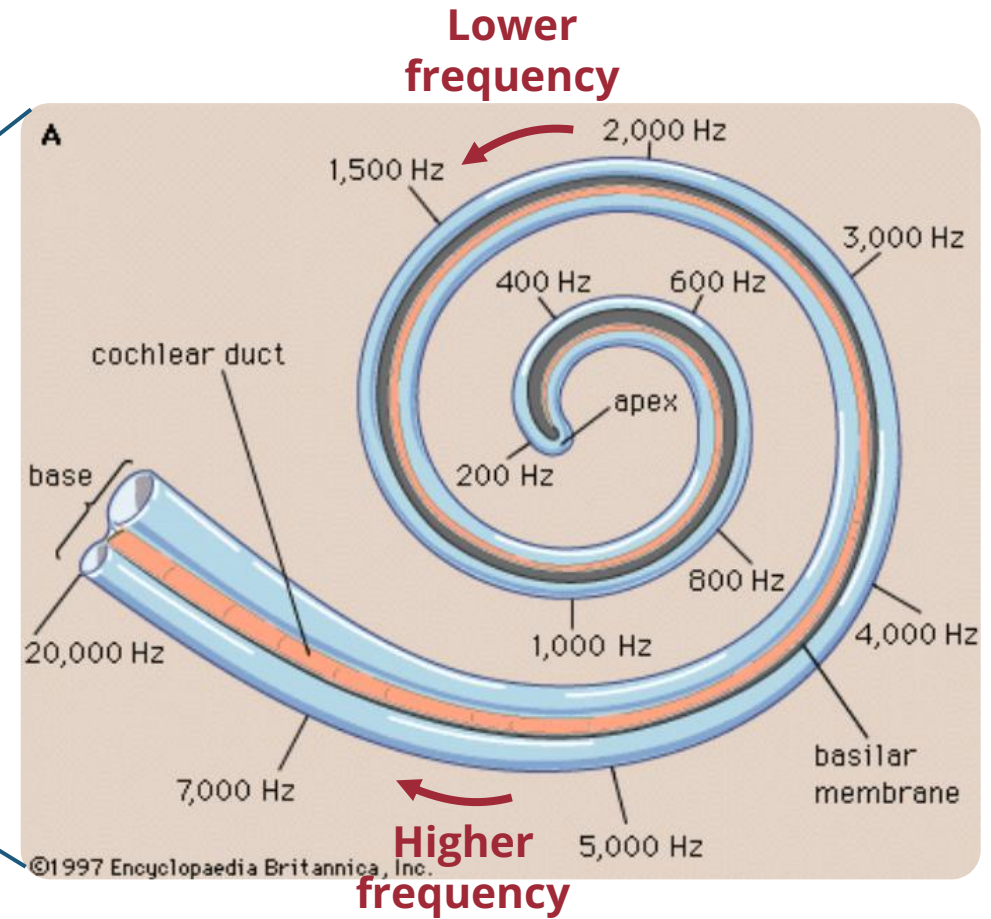


(Source: NIH/NIDCD)

Cochlea in the Inner Ear

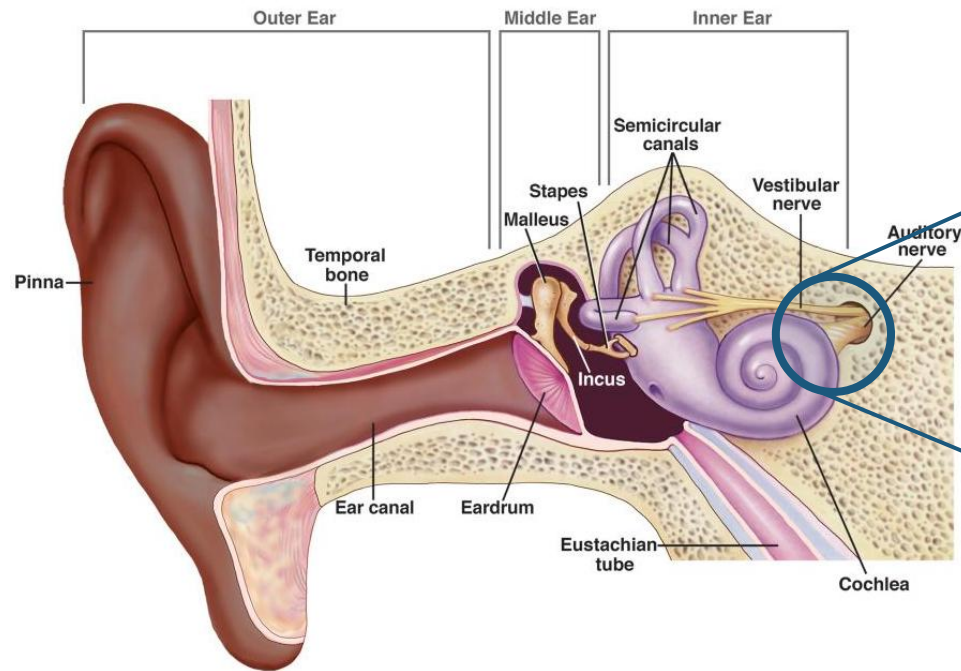


(Source: NIH/NIDCD)

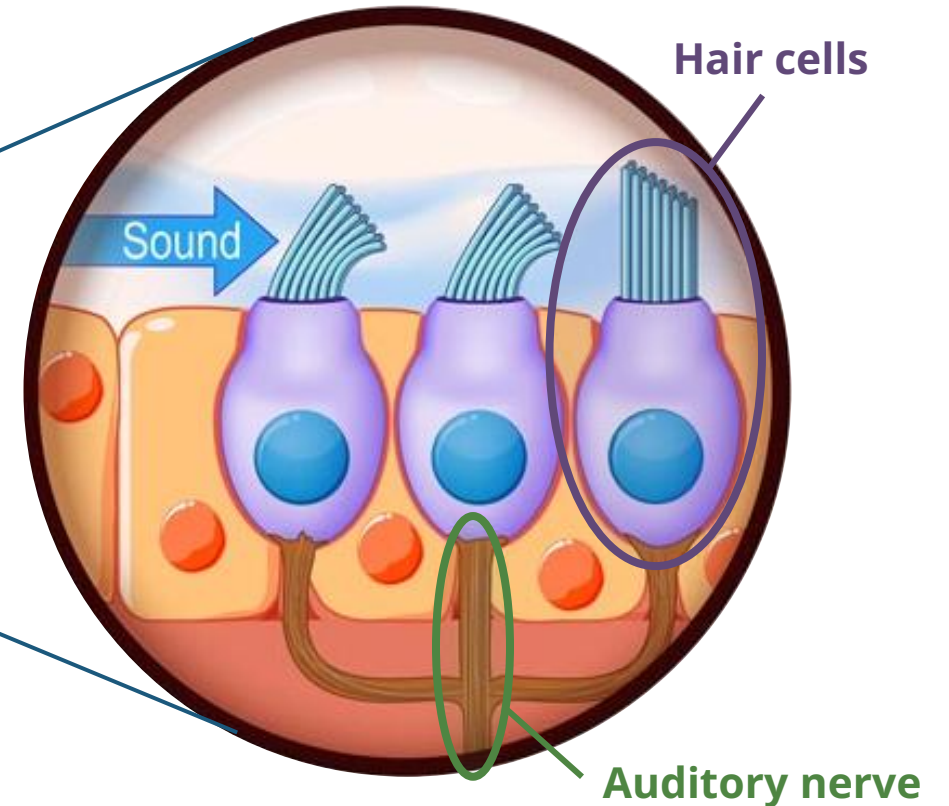


(Source: Britannica)

Hair Cells in the Cochlea

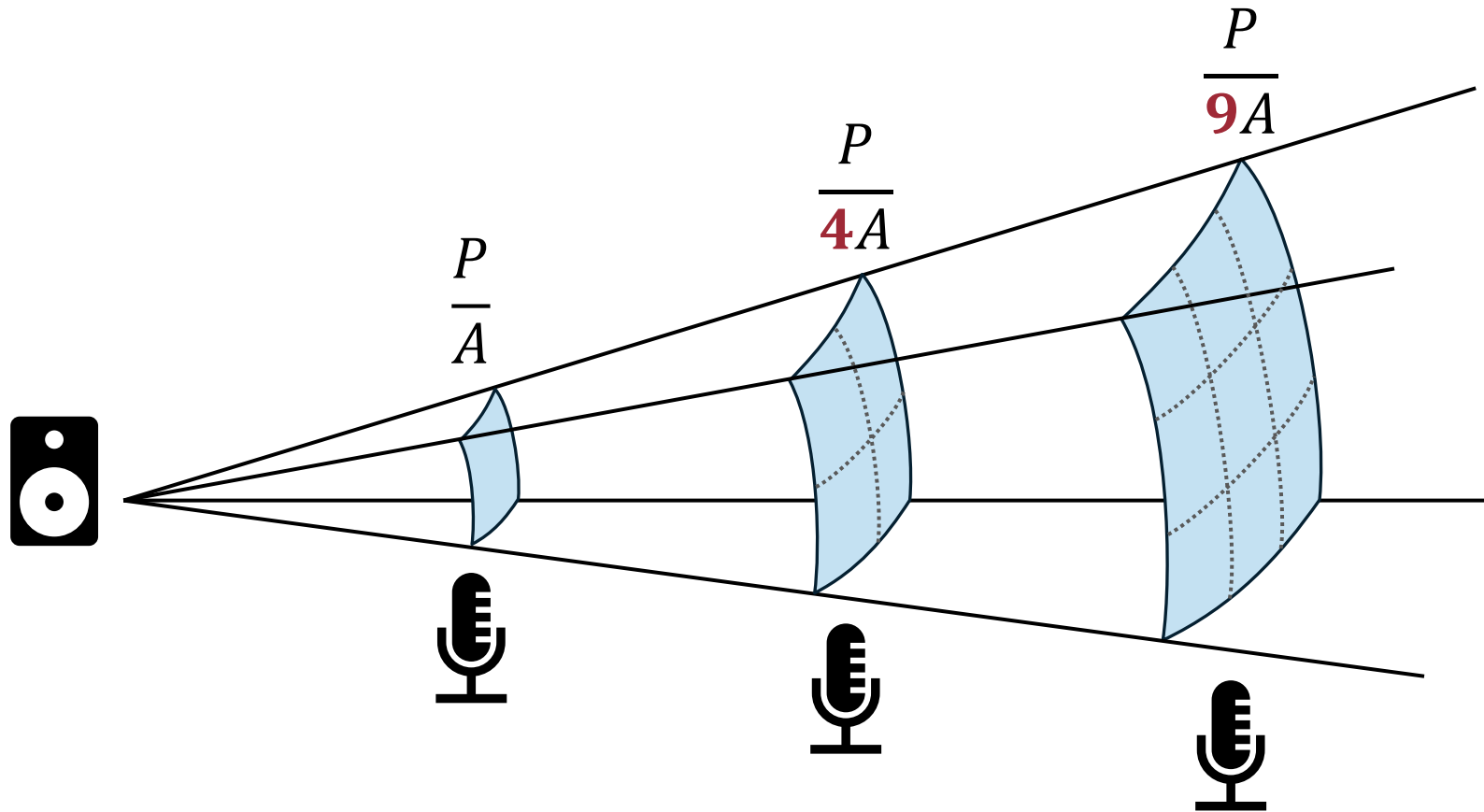


(Source: NIH/NIDCD)



(Source: COSMOS Magazine)

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

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

- $I_{\text{REF}} := 10^{-12} \text{W}/\text{m}^2$ 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)	

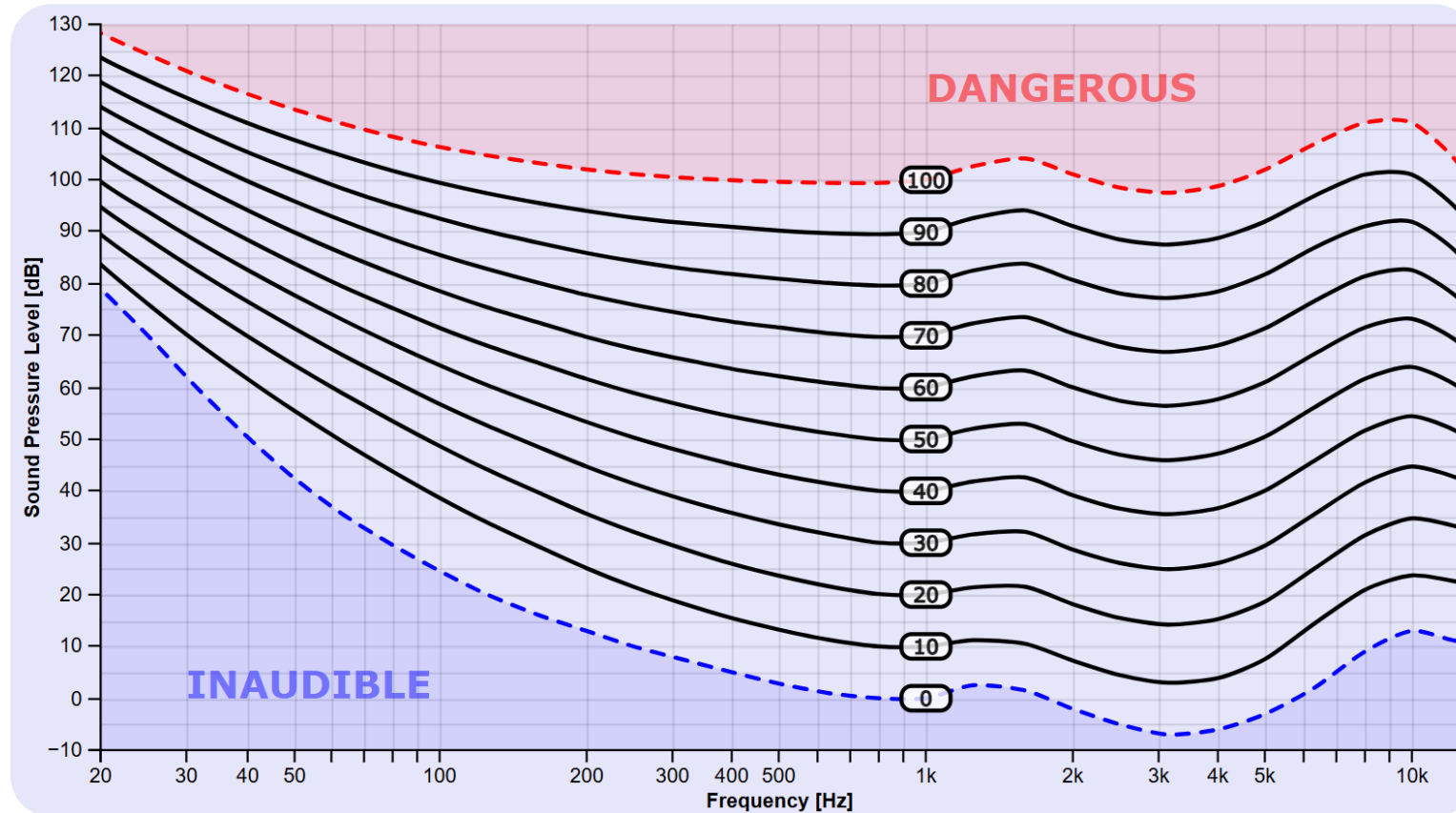
(Unit: W/m²)

(Source: Britannica)

🤔 Common Gains in Decibels

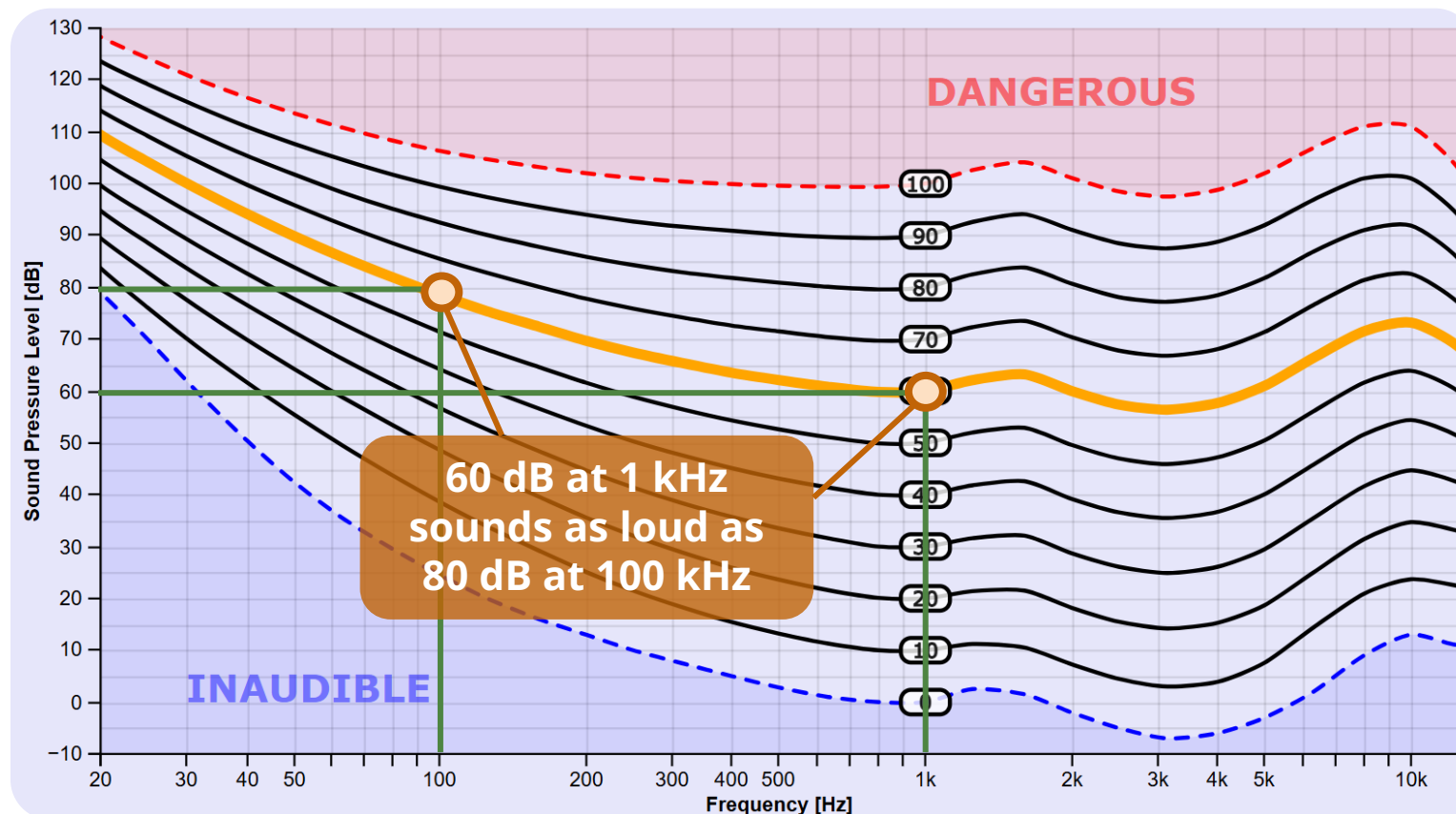
- **+10 dB = ?x intensity**
 - **10x intensity** (3.16x amplitude)
- **+3 dB ≈ ?x intensity**
 - **2x intensity** (1.414x amplitude)
- **+20 dB = ?x amplitude**
 - **10x amplitude** (100x intensity)
- **+6 dB ≈ ?x amplitude**
 - **2x amplitude** (4x intensity)

Loudness Perception: Equal-loudness Contours



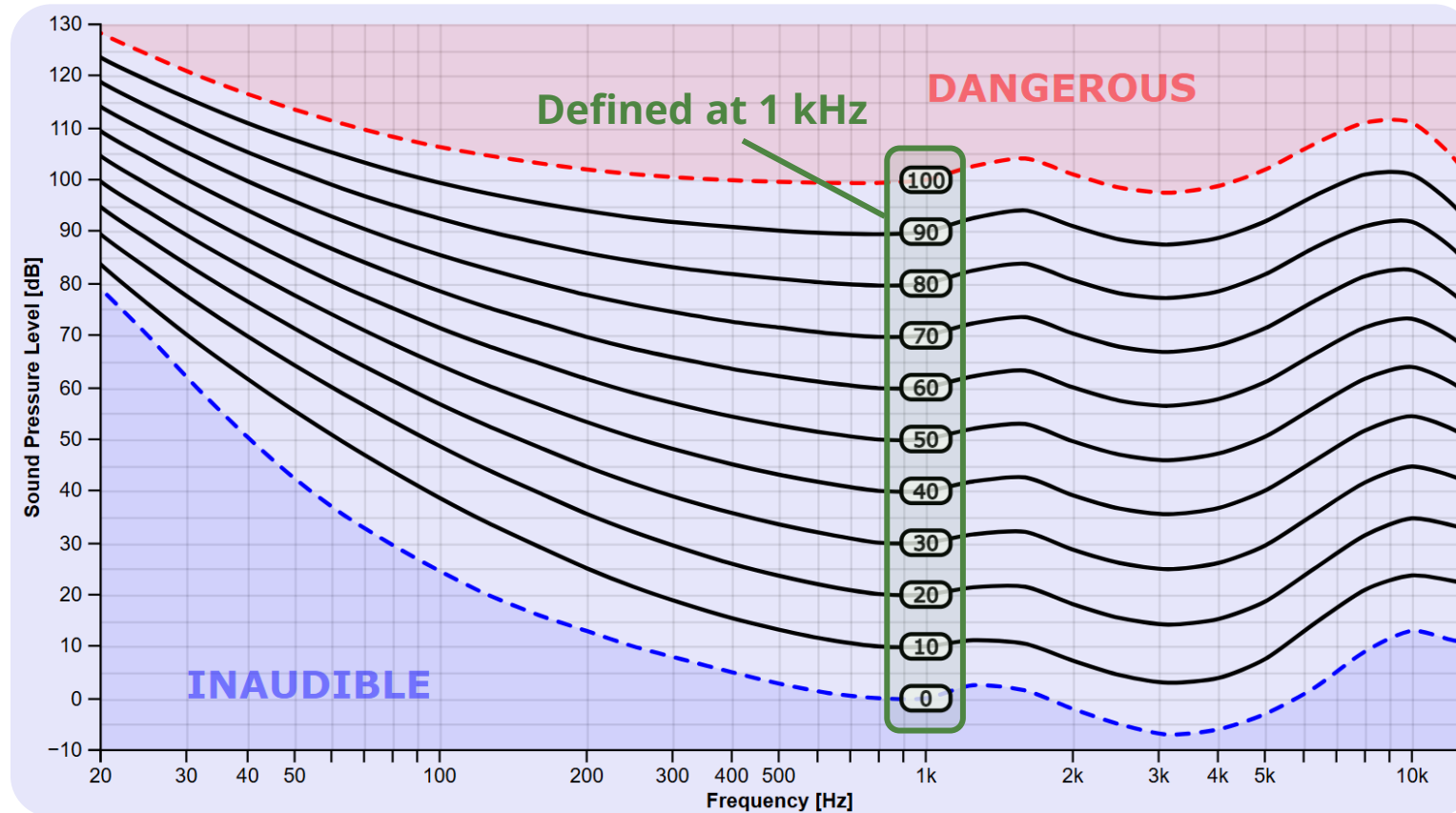
(Source: Parker, 2024)

Loudness Perception: Equal-loudness Contours



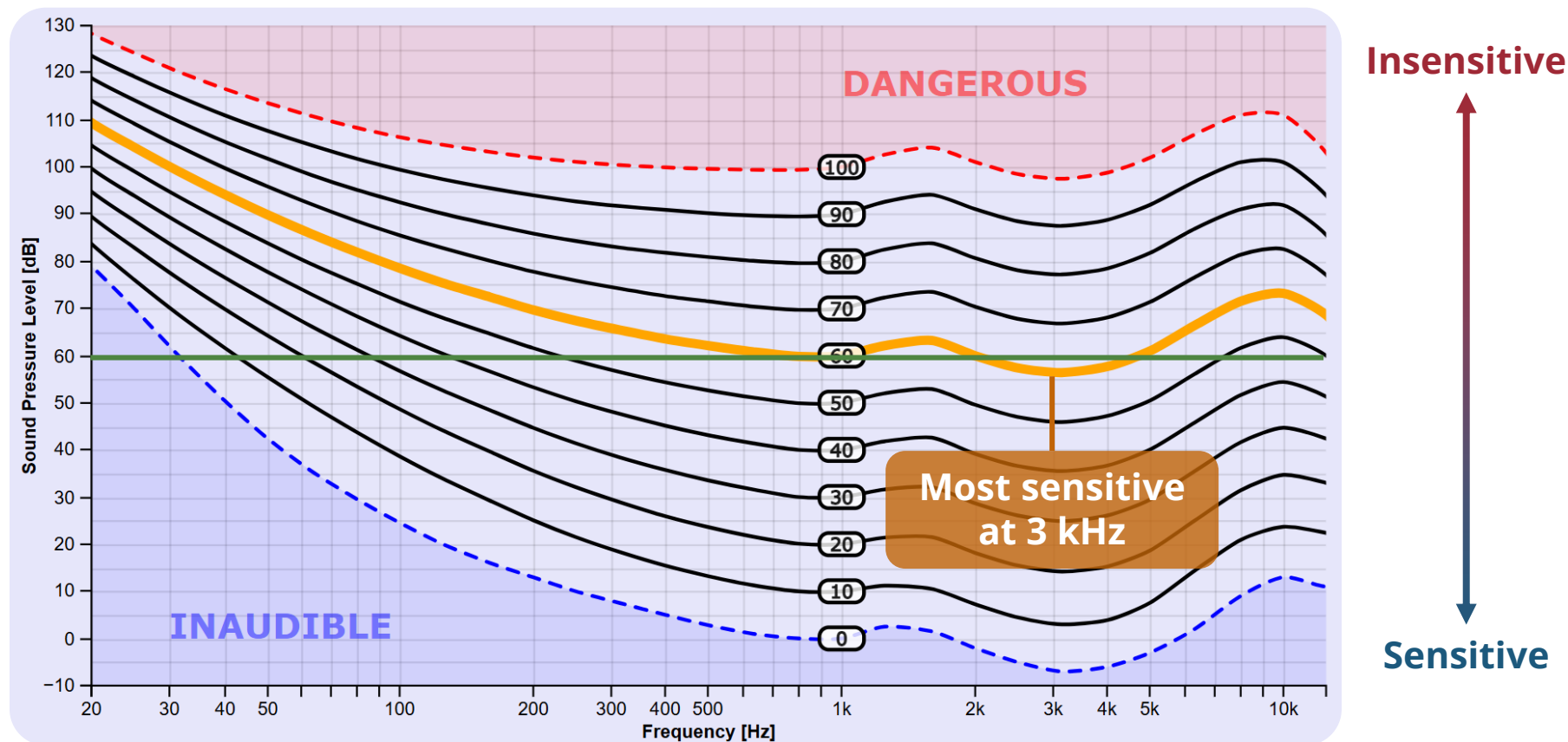
(Source: Parker, 2024)

Loudness Perception: Equal-loudness Contours



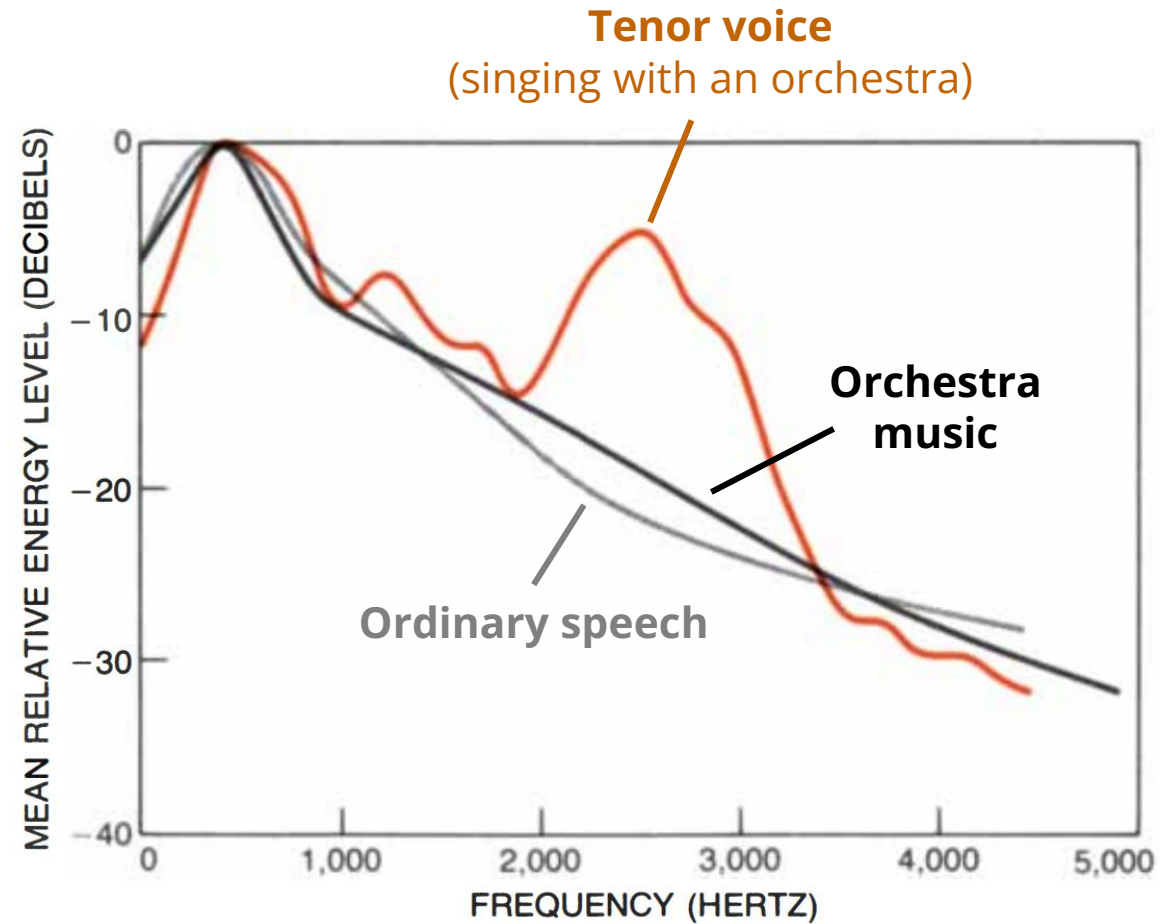
(Source: Parker, 2024)

Loudness Perception: Equal-loudness Contours



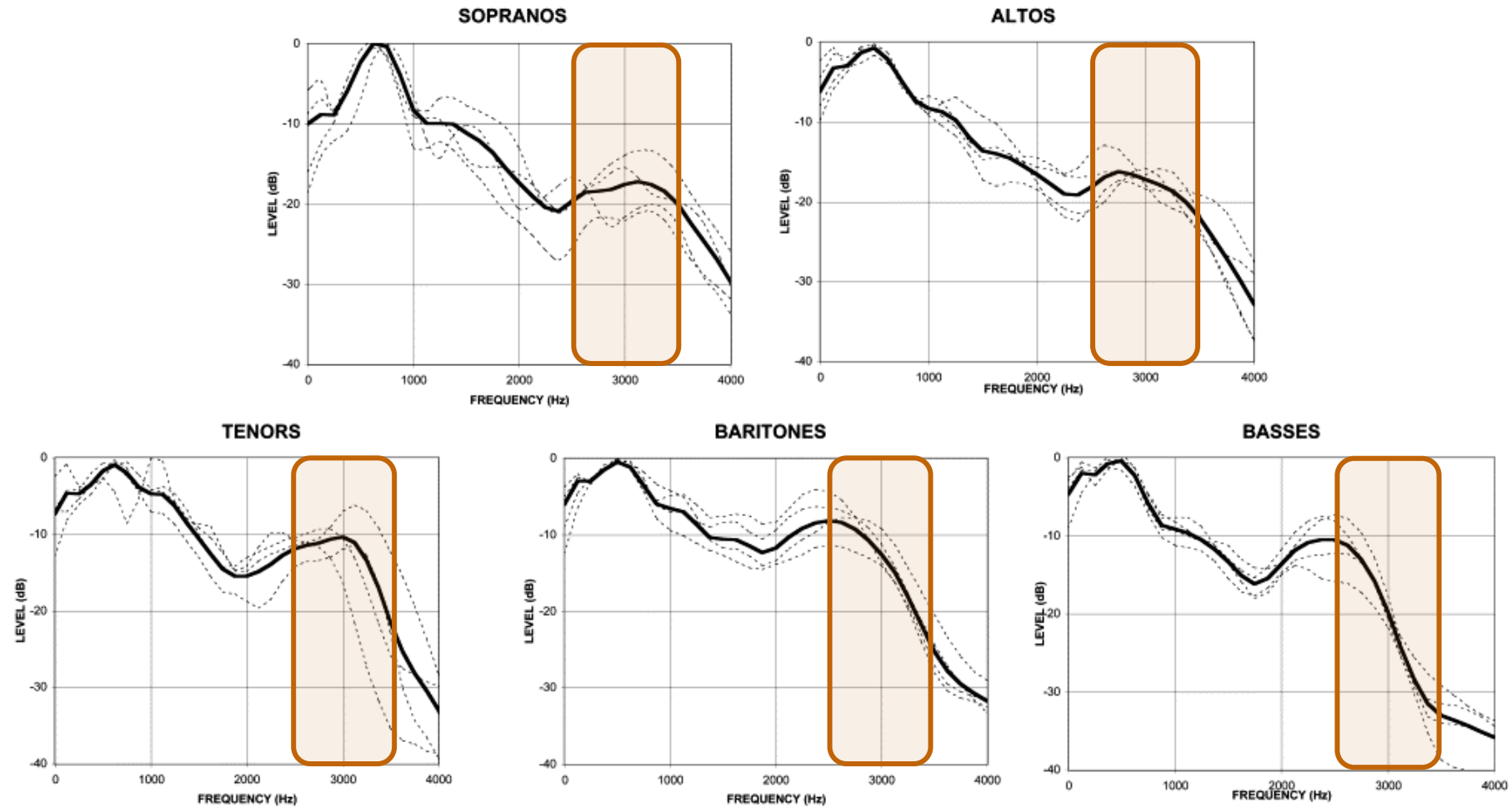
(Source: Parker, 2024)

Singer's Formants (Sundberg, 1991)



(Source: Sundberg, 1977)

Singer's Formants (Sundberg, 1991)



(Source: Sundberg, 1977)

| Psychoacoustics

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

Recap

Four Representative Music Representations



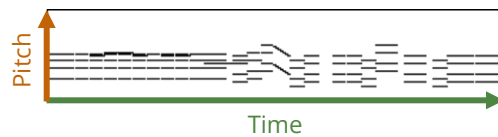
Symbolic music representations

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



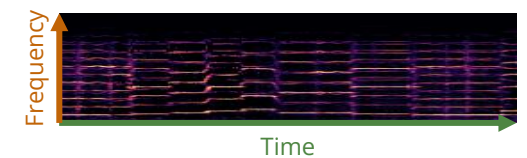
Audio-domain music representations

Time series-based



Waveform

Image-based



Spectrogram

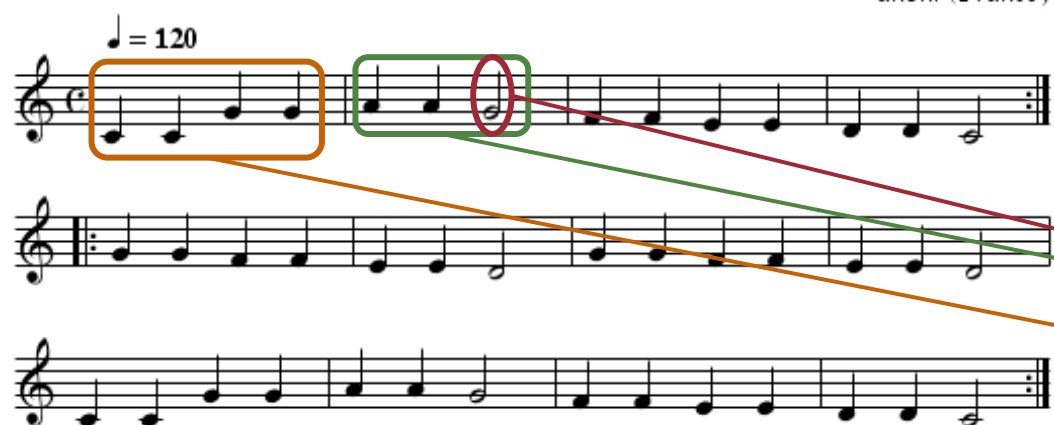
Today's topic!

An Example of ABC Notation

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

anon. (France)

$\text{♩} = 120$

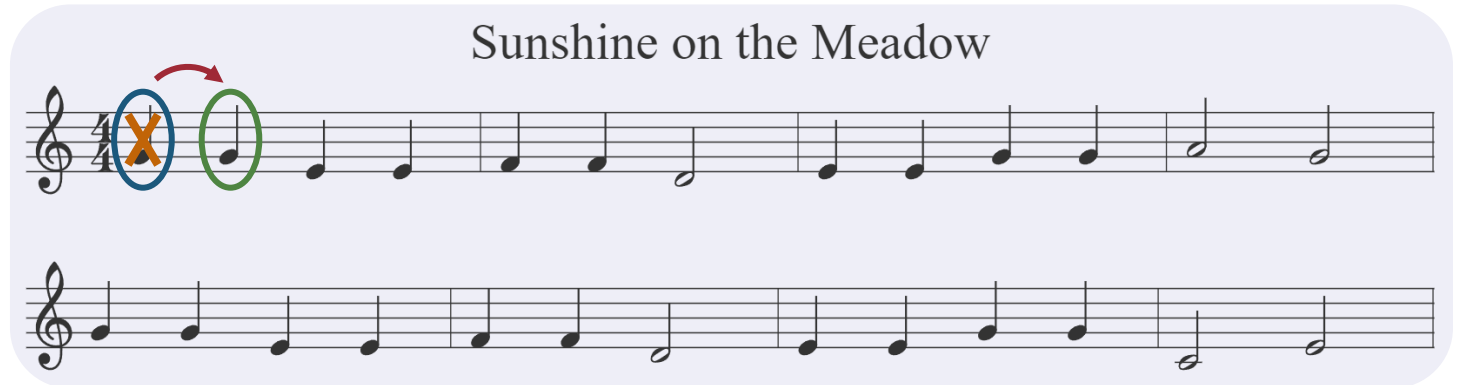


Metadata

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: |

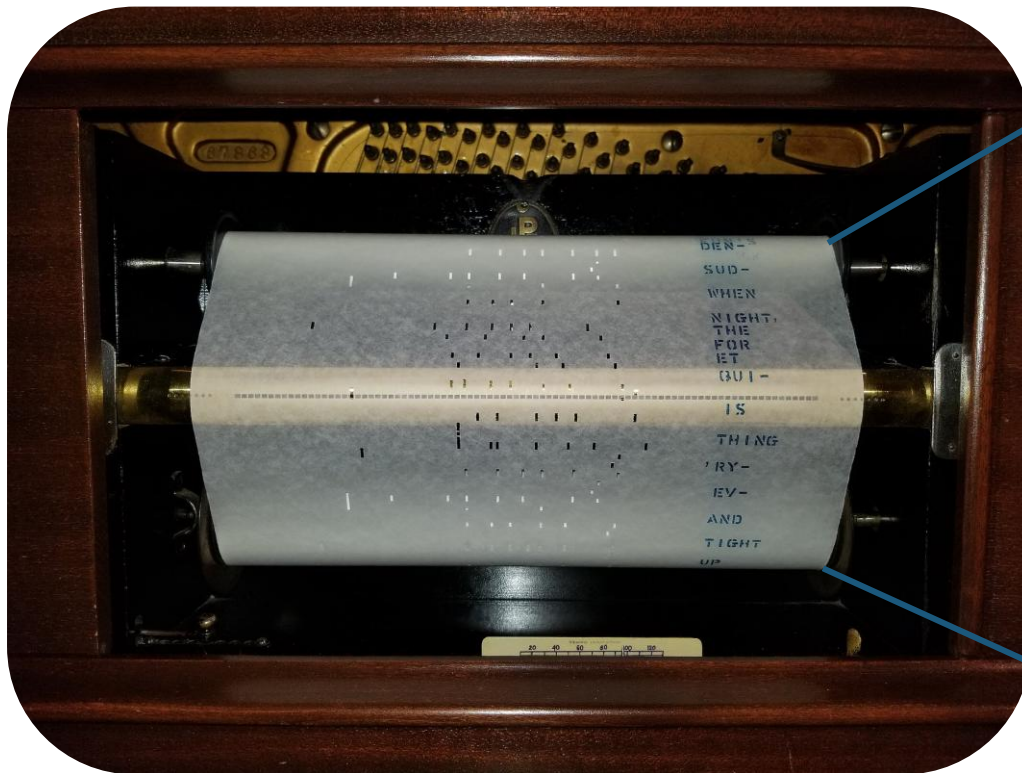
Representing Music using MIDI Messages

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

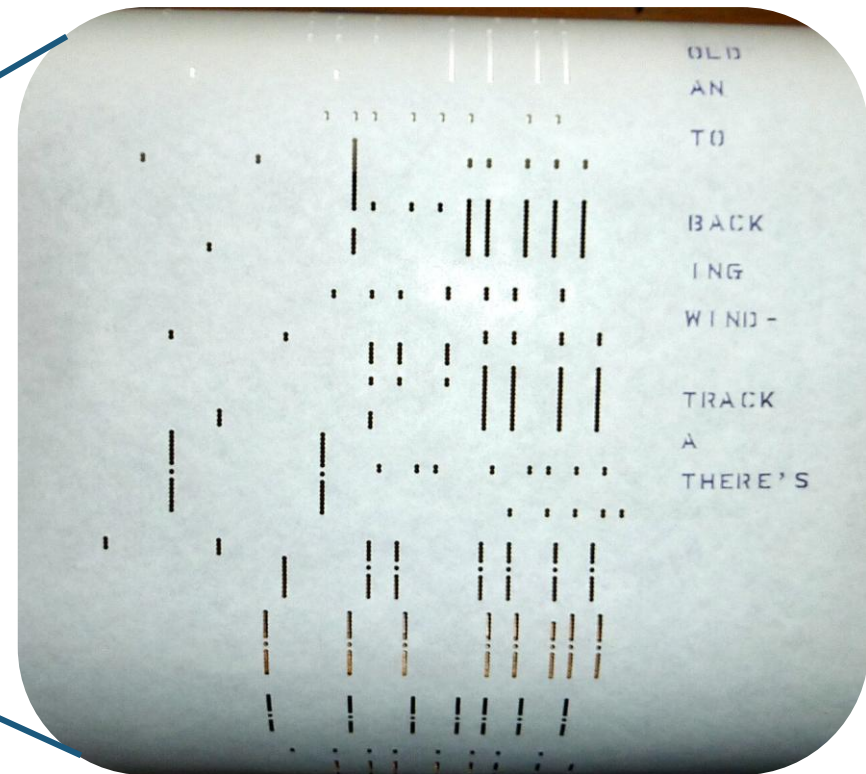


```
Note_on_67, Time_shift_quarter_note, Note_off_67,  
Note_on_67, Time_shift_quarter_note, Note_off_67,  
Note_on_64, Time_shift_quarter_note, Note_off_64,  
Note_on_64, Time_shift_quarter_note, Note_off_64,  
...
```

Piano Rolls

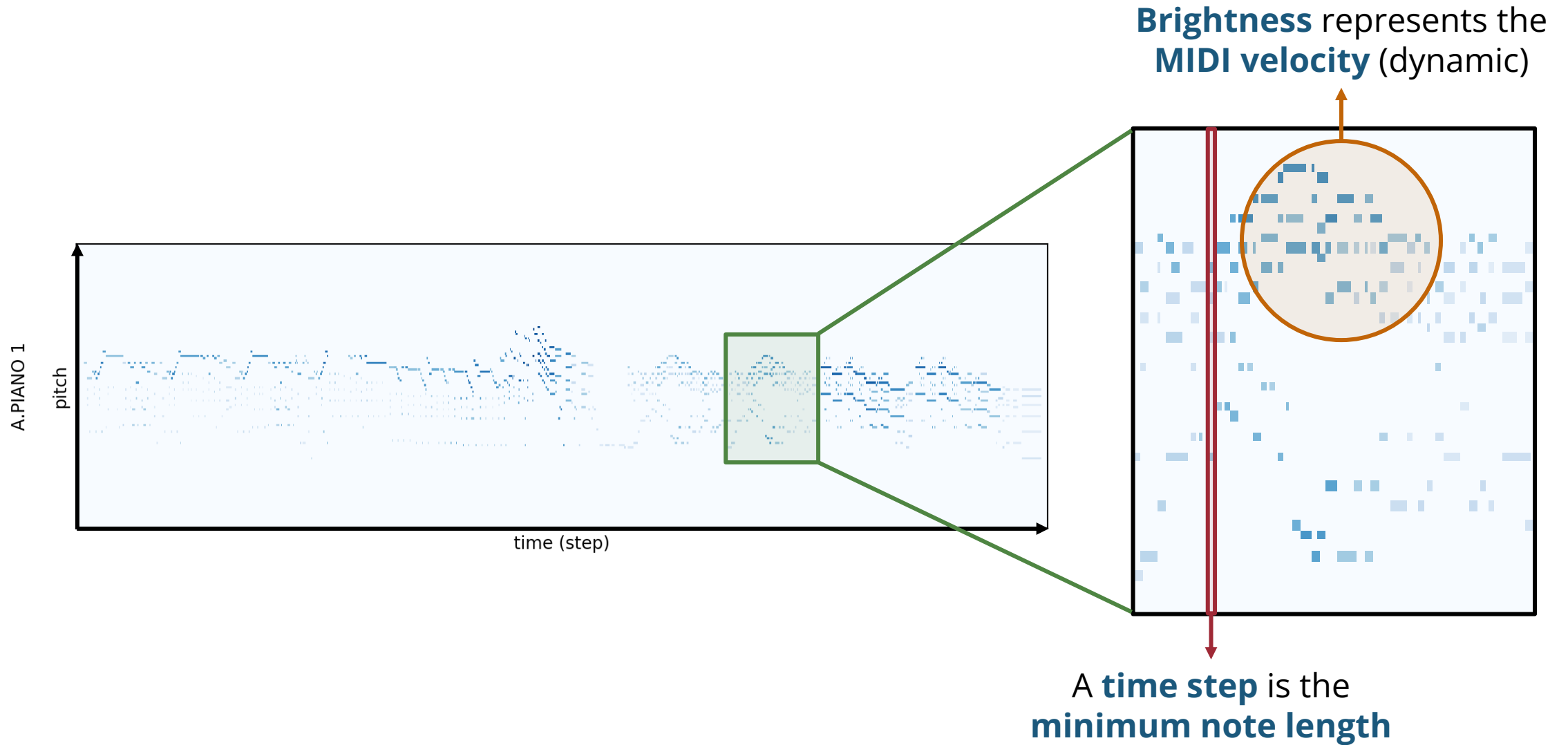


(Source: Draconichiaro)

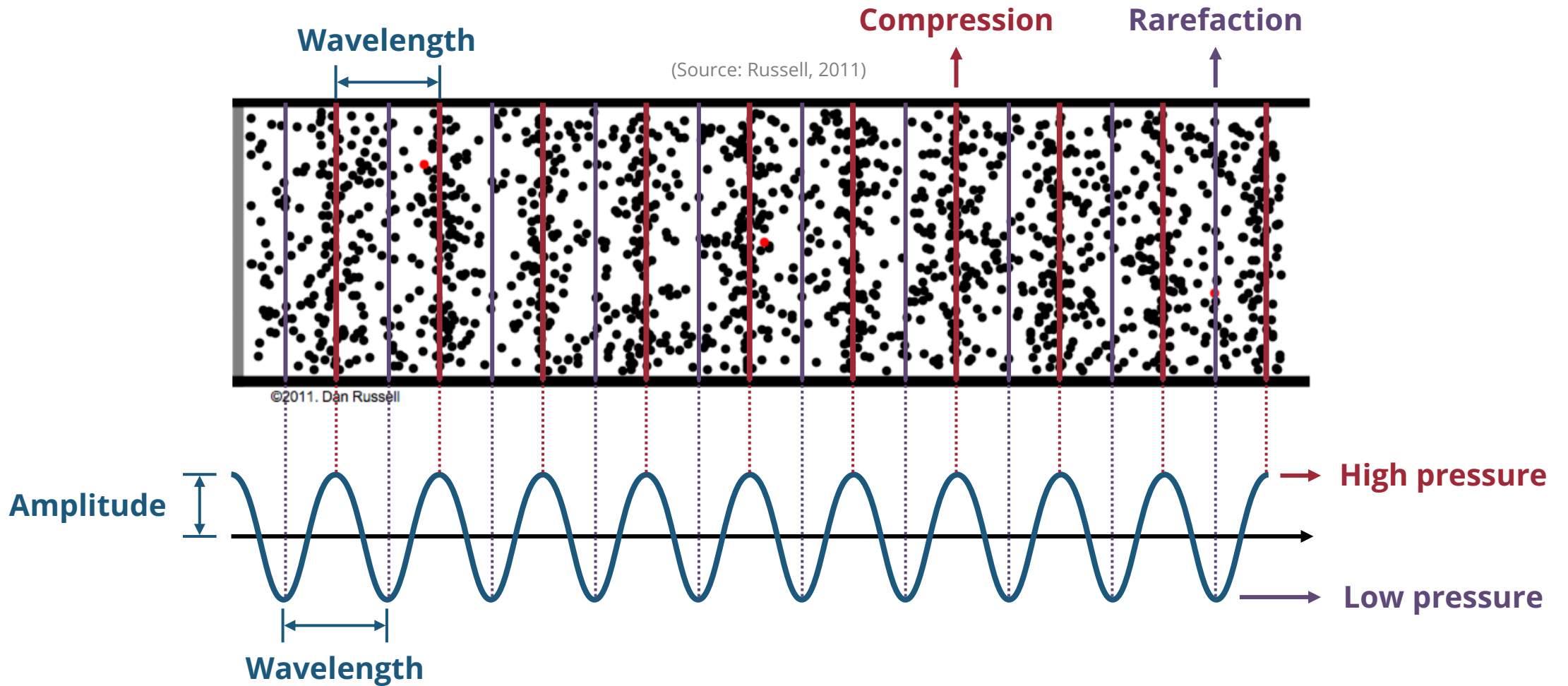


(Source: Tangerineduel)

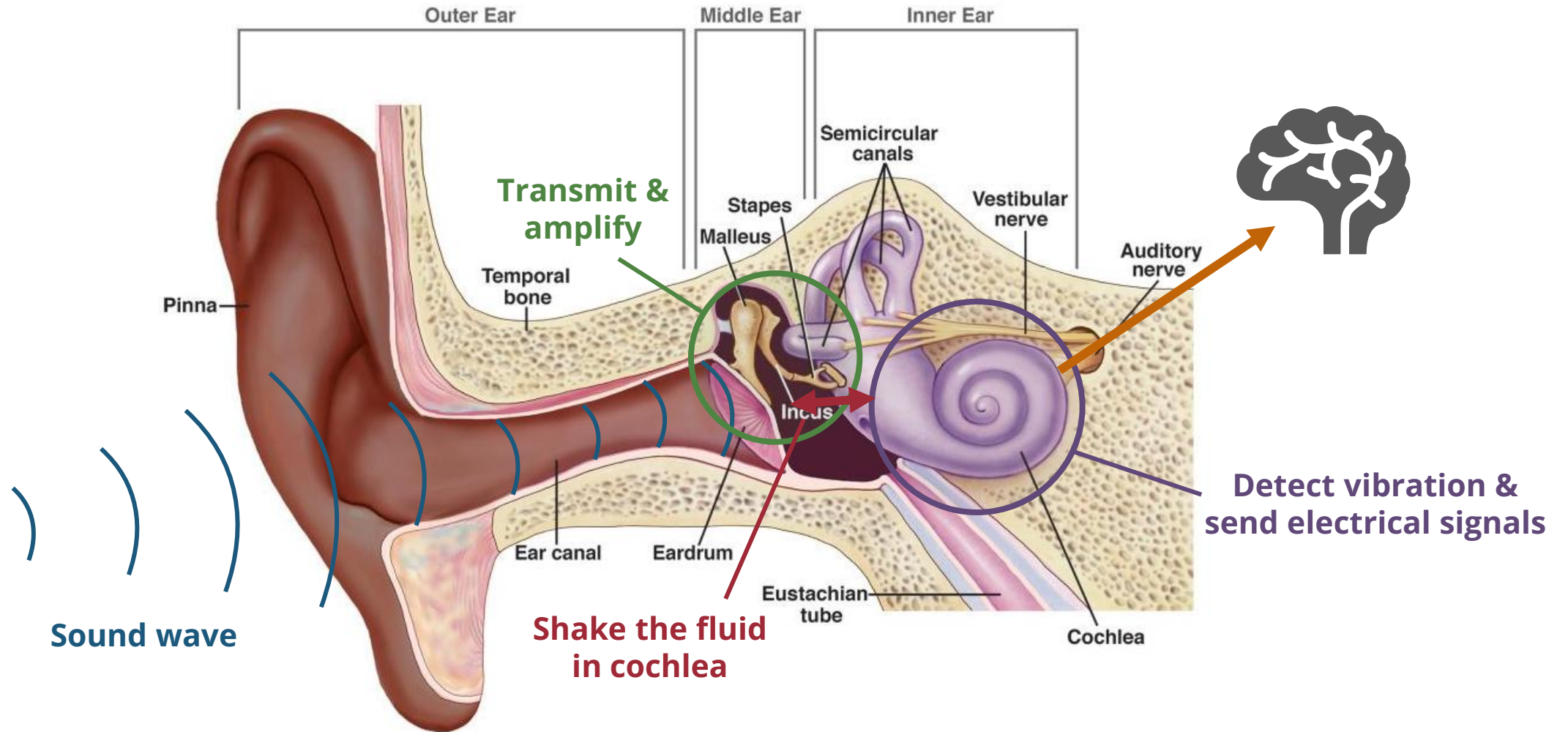
Piano Roll Representation



Longitudinal vs Transverse Waves

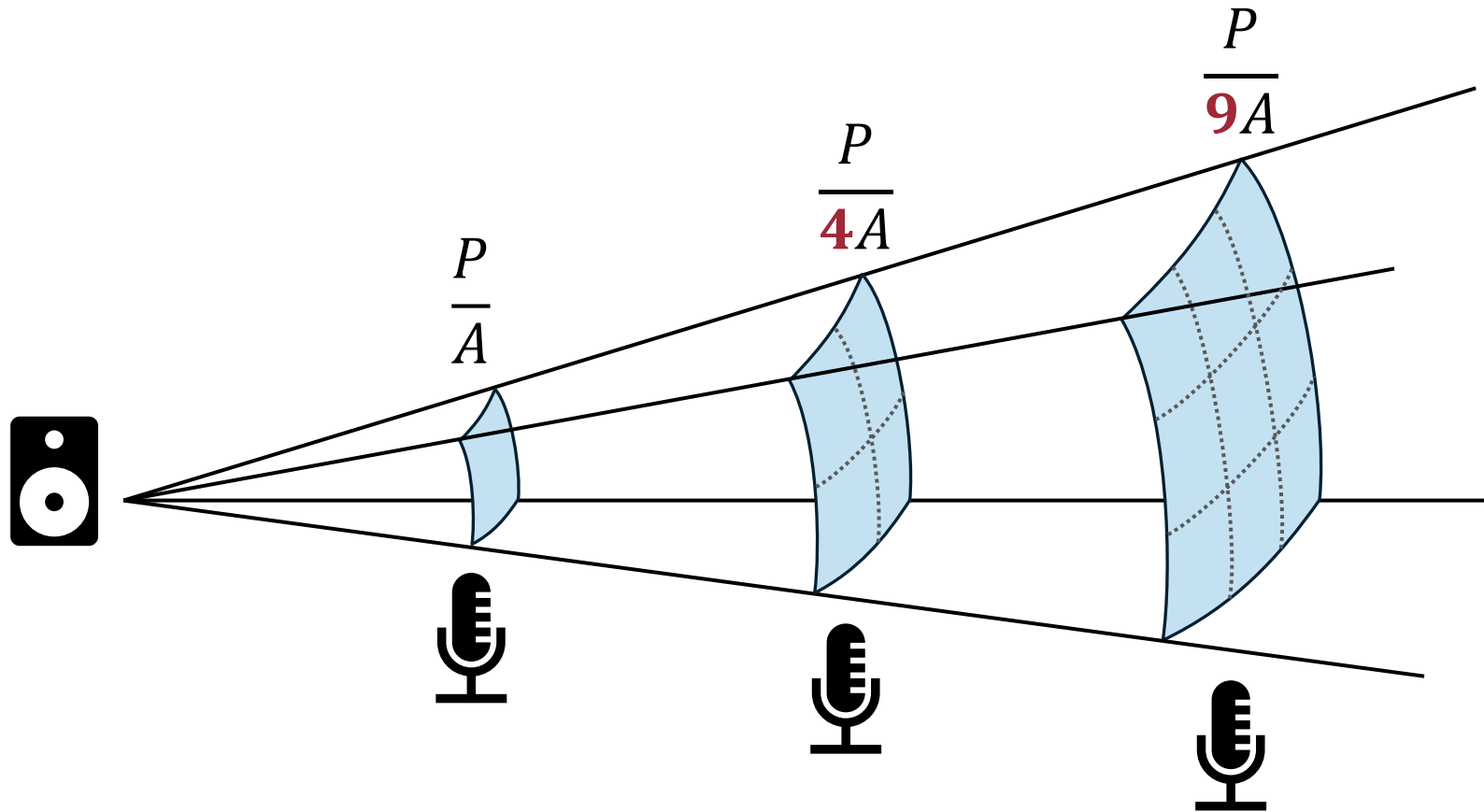


Human Ears



(Source: NIH/NIDCD)

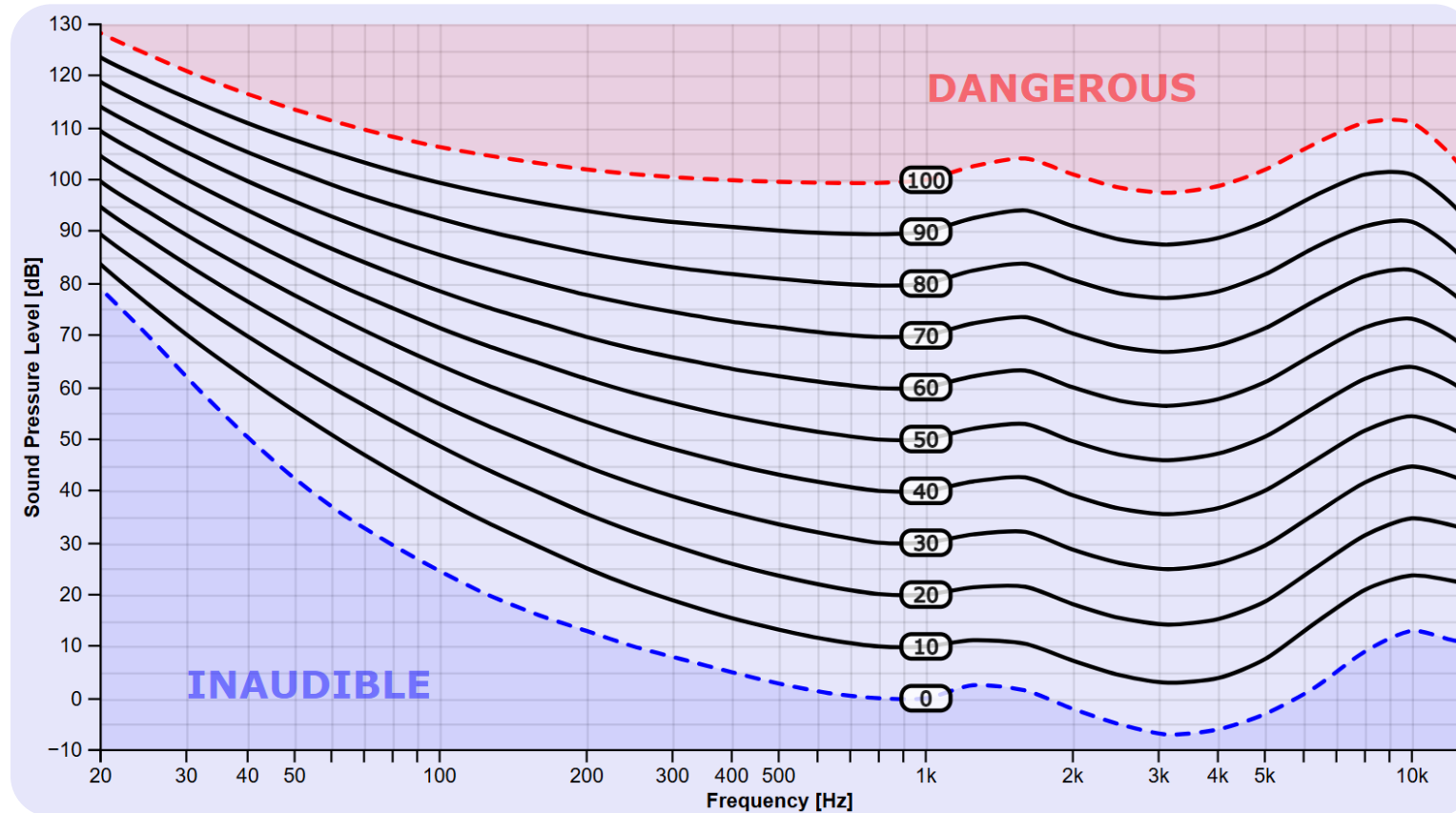
Sound Propagation & Inverse Square Law



Inverse Square Law

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

Loudness Perception: Equal-loudness Contours

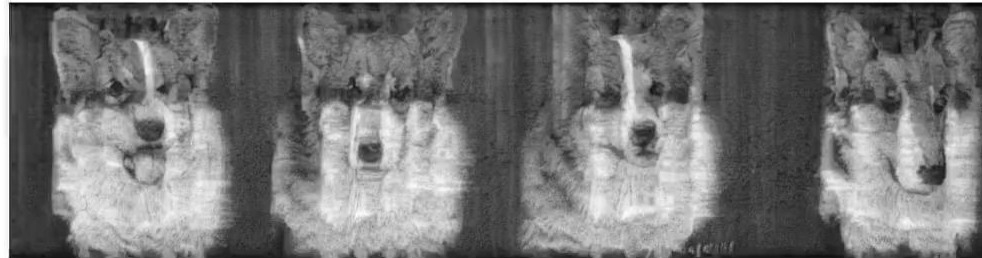
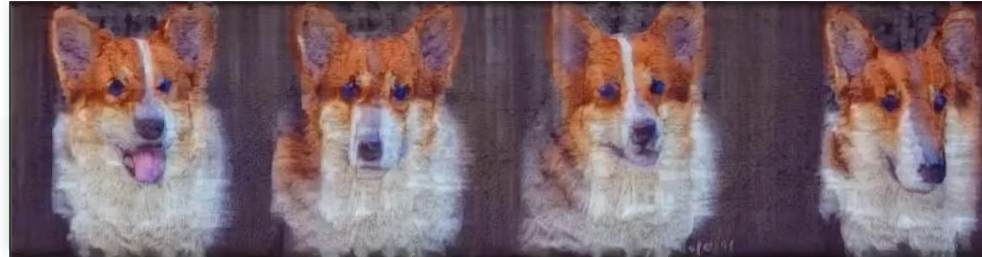


(Source: Parker, 2024)

Next Lecture

Audio Processing Fundamentals

Image prompt: a colorful photo of corgis



Audio prompt: dog barking

(Source: Chen et al., 2024)