Why is AI so Bad at Music?

(at least so far)?

Chris White
Dept. of Music and Dance
UMass Amherst

UNIVERSITY OF MASSACHUSETTS
1863

HFA College of Humanities & Fine Arts

There's a lot of Artificial Intelligence swirling around us

New Tool for Building and Fixing Roads and Bridges: Artificial Intelligence

In Pennsylvania and elsewhere, A.I. is being applied to the nation's aging infrastructure. Is that wise?

What Do You Do When A.I. Takes Your Voice?

Two voice actors say an A.I. company created clones of their voices without their permission. Now they're suing. The company denies it did anything wrong.

Can Artificial Intelligence Make the PC Cool Again?

An A.I. Robot Named Sophia Tells Graduates to Believe in Themselves

D'Youville University in Buffalo had an A.I. robot speak at its commencement on Saturday. Not everyone was happy about it.

Can Google Give A.I. Answers Without Breaking the Web?

Publishers have long worried that artificial intelligence would drive readers away from their sites. They're about to find out if those fears are warranted.

A.I.'s 'Her' Era Has Arrived

New chatbot technology can talk, laugh and sing like a human. What comes next is anyone's guess.

There's a lot of Artificial Intelligence swirling around us

AI

Udio and Suno lead the battle of the AI music generators

Udio raised $10 million in funding from high-profile investors like a16z, will.i.am, Common, and Instagram co-founder Mike Krieger.

The AI Music Era Is Here. Not Everyone Is a Fan

AI songwriting has gotten shockingly good — with big implications for the music world.

…but you have to look much harder to find buzz about AI that actually makes music

Here's what it sounds like

AI

Udio and Suno lead the battle of the AI music generators

Udio raised $10 million in funding from high-profile investors like a16z, will.i.am, Common, and Instagram co-founder Mike Krieger.

The AI Music Era Is Here. Not Everyone Is a Fan

AI songwriting has gotten shockingly good — with big implications for the music world.

Music is hard for Large Language Models

- Musical AI is technologically behind
- Its output is less convincing
- It gets less attention from media and users
- So… why is music so hard for for LLMs?

Five Forces Effecting Musical AI

Motivation

Why are people making models of musical AI?

Five Forces Effecting Musical AI

Motivation

Why are people making models of musical AI?

Examples

What datasets are people using to train and test their AI models?

Five Forces Effecting Musical AI

Motivation	**E**xamples	**R**epresentation
Why are people making models of musical AI?	What datasets are people using to train and test their AI models?	How are programmers representing musical events in their AI?

Five Forces Effecting Musical AI

Motivation	**Examples**	**Representation**	**Structure**
Why are people making models of musical AI?	What datasets are people using to train and test their AI models?	How are programmers representing musical events in their AI?	What aspects of musical organization are being learned by the AI?

Five Forces Effecting Musical AI

Motivation	Examples	Representation	Structure	Interpretation
Why are people making models of musical AI?	What datasets are people using to train and test their AI models?	How are programmers representing musical events in their AI?	What aspects of musical organization are being learned by the AI?	What value are listeners drawing from music generated by AI?

Five Forces Effecting Musical AI

…and they each have downstream effects on the next

First some definitions

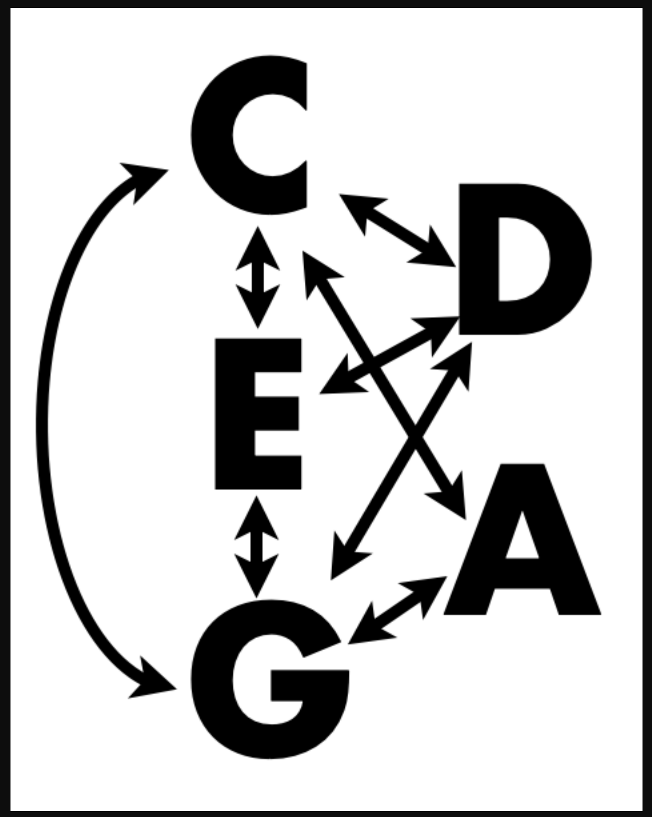
- Machine Learning: A computer system that learns something about some medium or topic by observing a relevant dataset

First some definitions

- Machine Learning: A computer system that learns something about some medium or topic by observing a relevant dataset

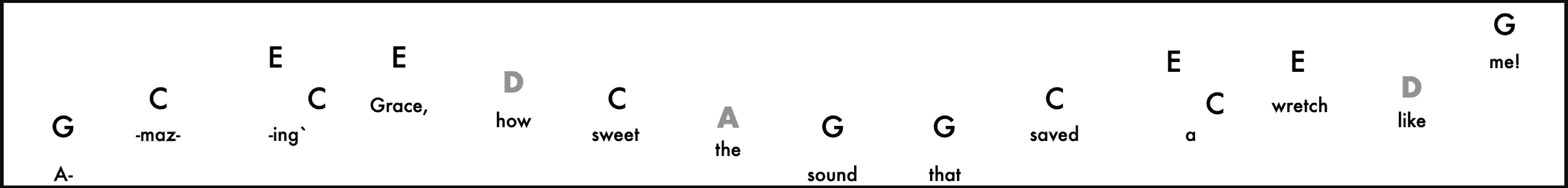
First some definitions

- Machine Learning: A computer system that learns something about some medium or topic by observing a relevant dataset



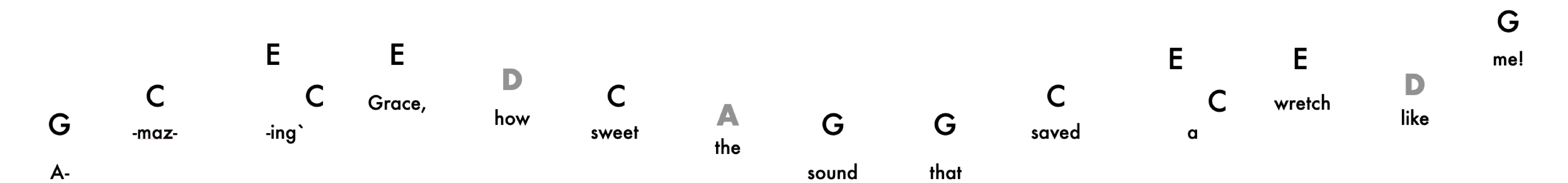
First some definitions

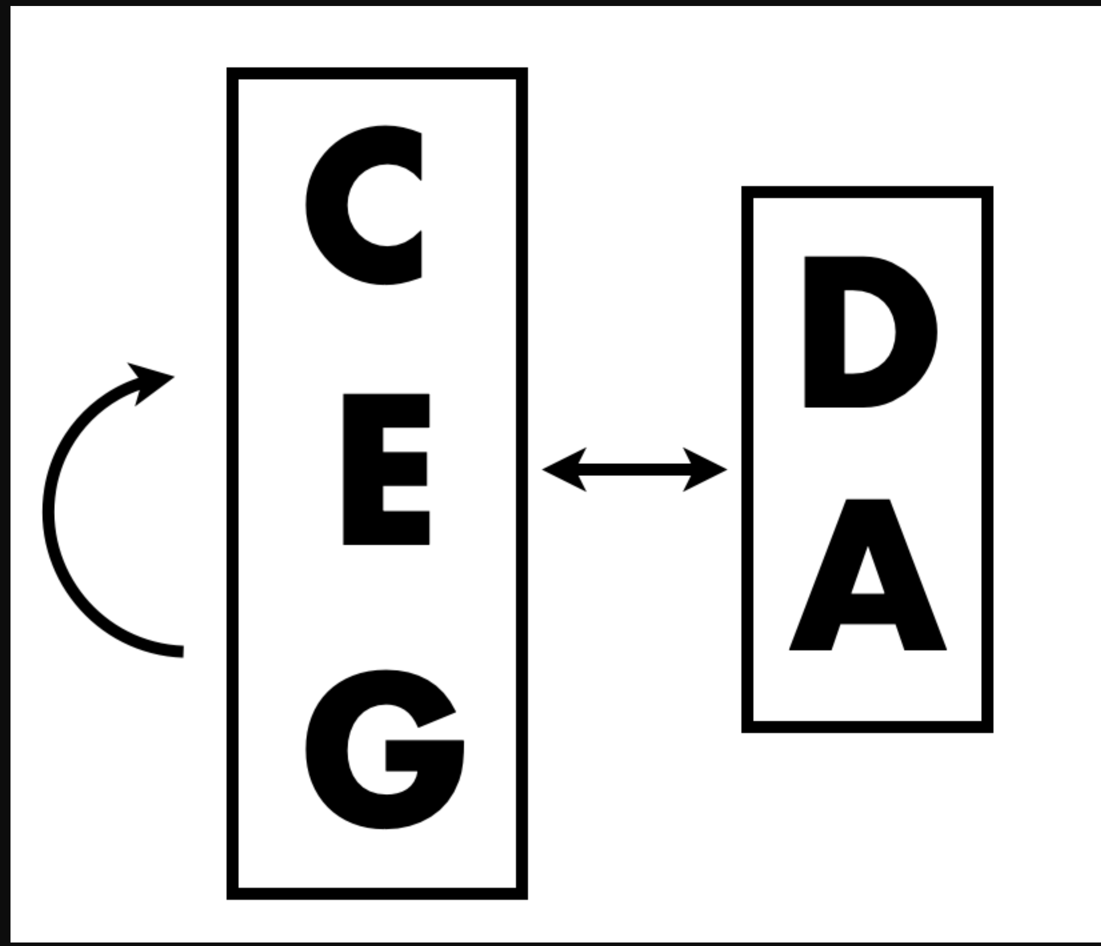
- Deep Learning: A computer system that learns by grouping observed events. Computationally, observed events become connected through deeper layers or categories



First some definitions

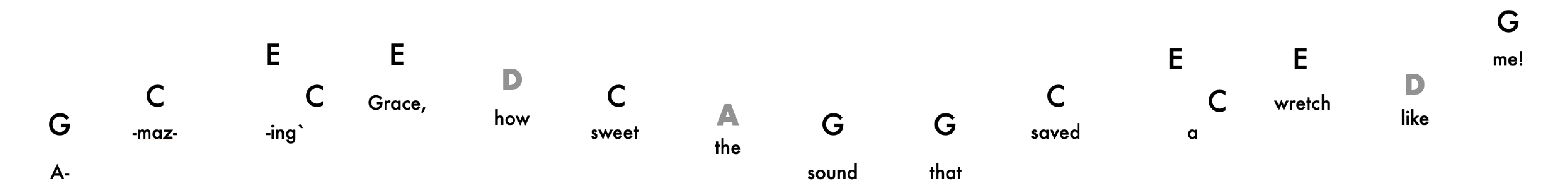
- Deep Learning: A computer system that learns by grouping observed events. Computationally, observed events become connected through deeper layers or categories

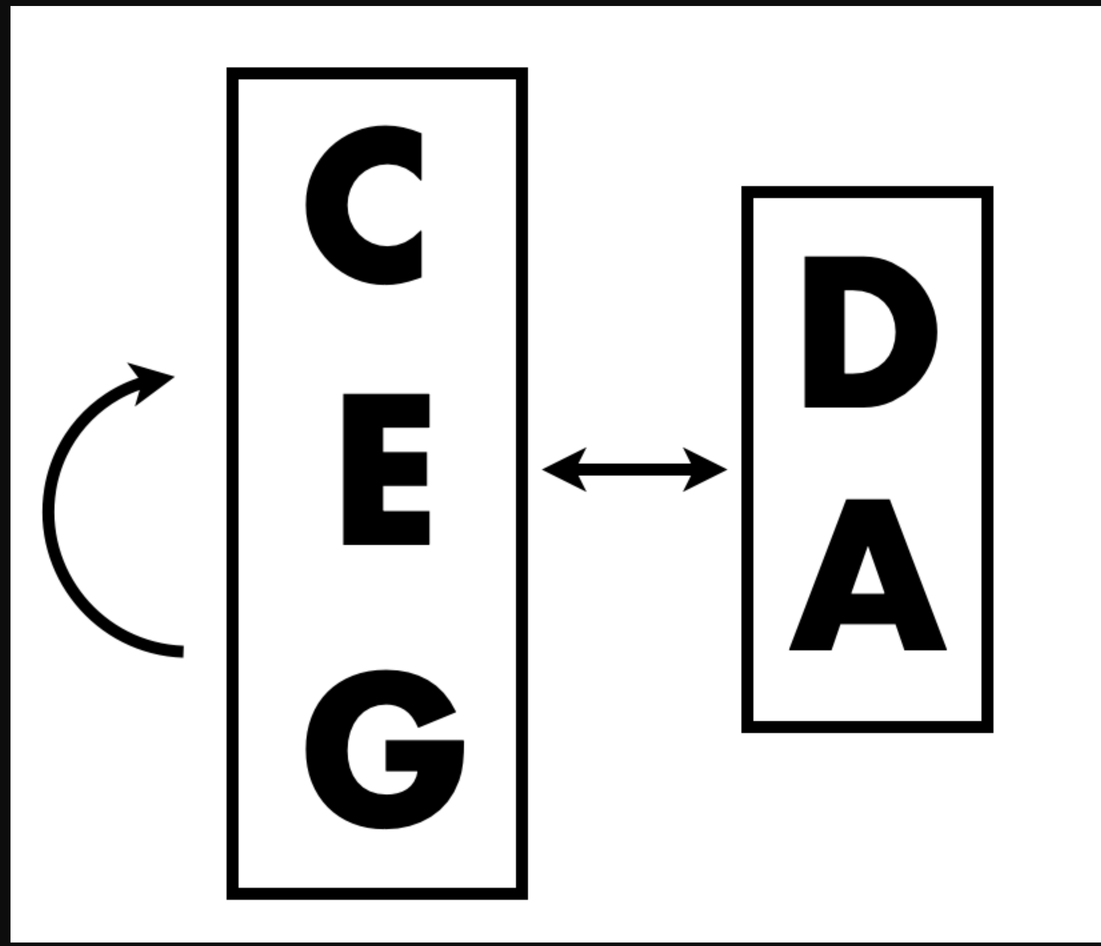


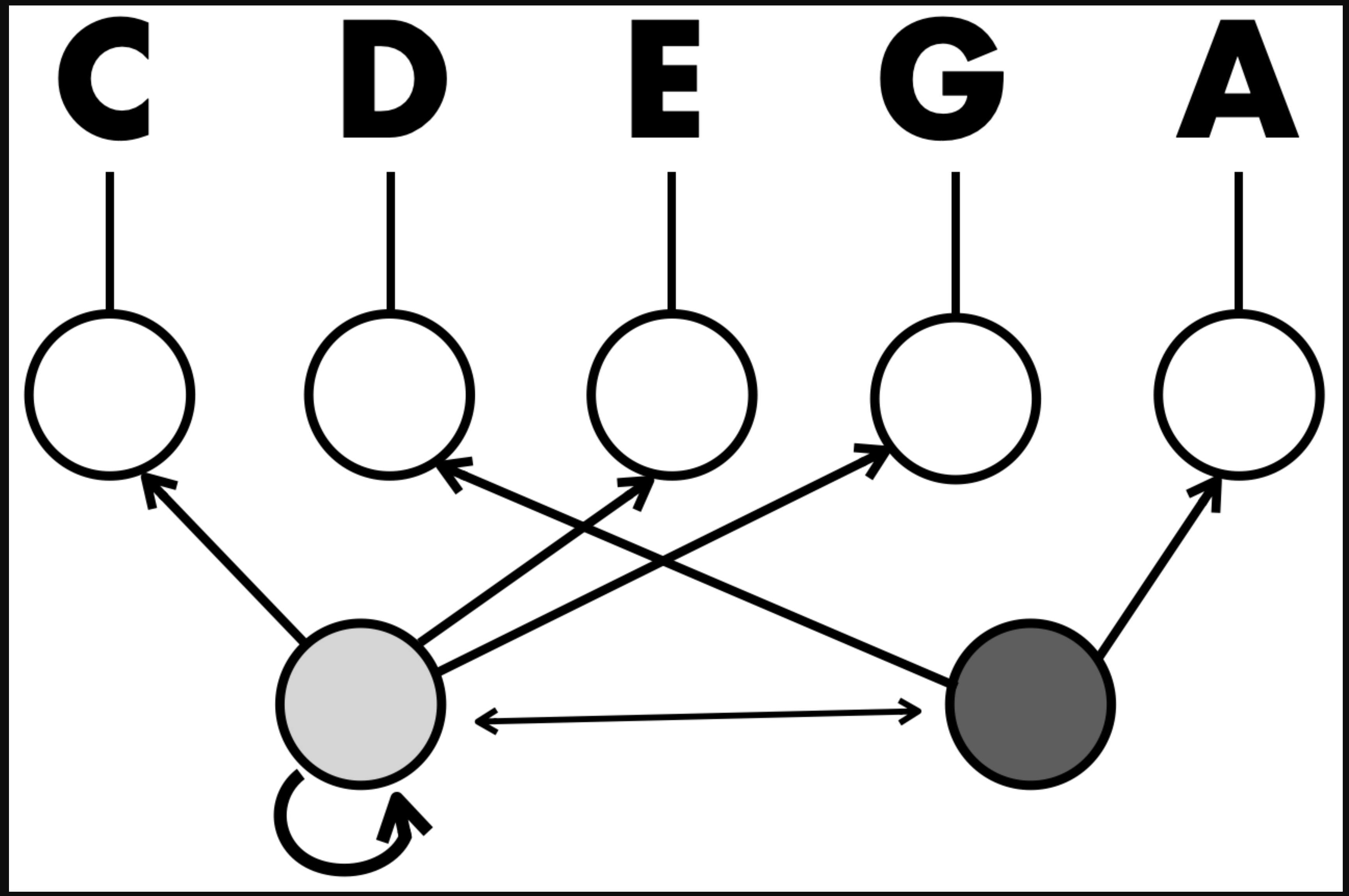


First some definitions

- Deep Learning: A computer system that learns by grouping observed events. Computationally, observed events become connected through deeper layers or categories







First some definitions

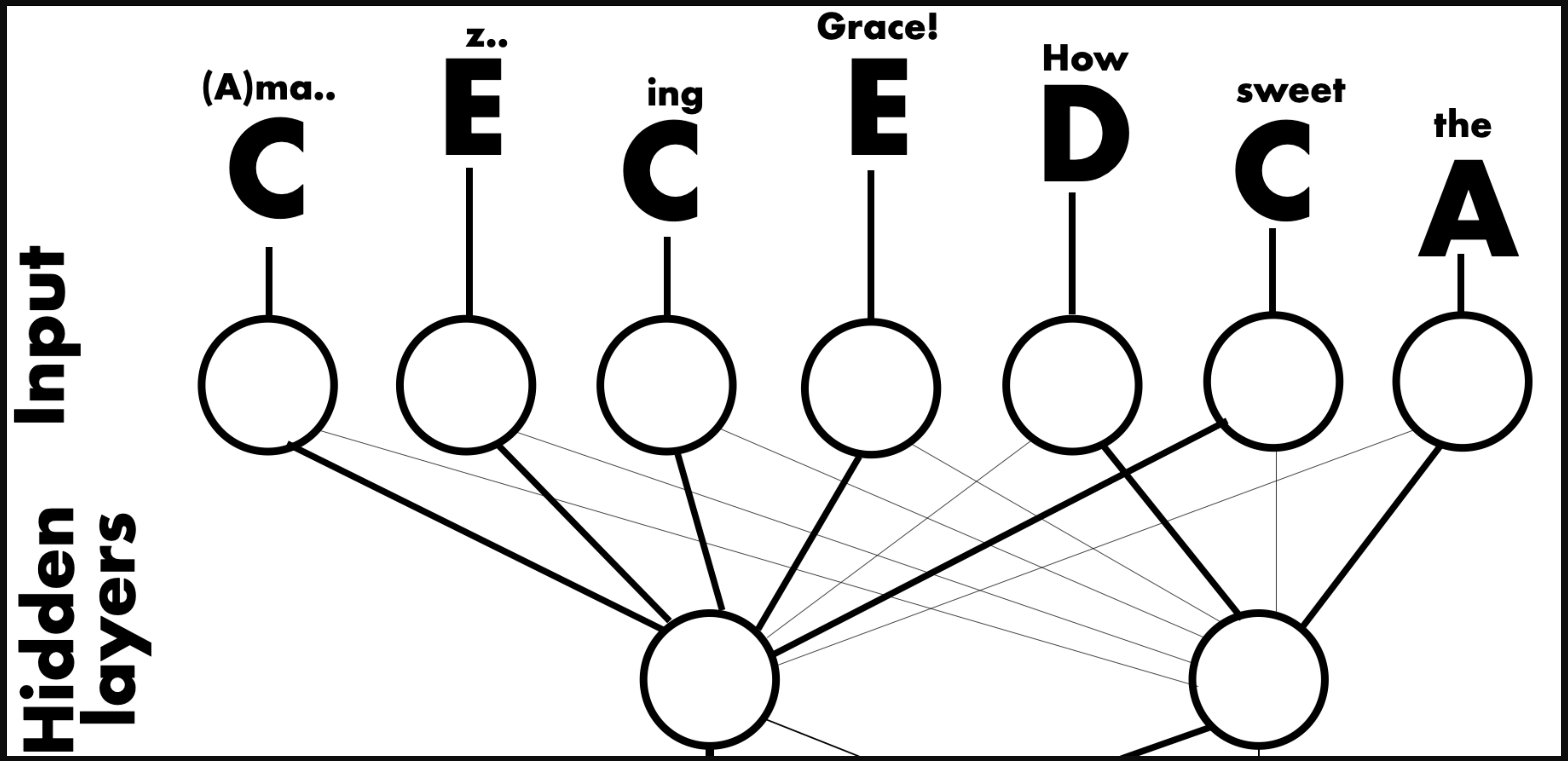
- Neural Network: A deep-learning system that uses multiple layers to capture complex connections between some observed input and output

- A transformer is a type of neural network

First some definitions

- Neural Network: A deep-learning system that uses multiple layers to capture complex connections between some observed input and output

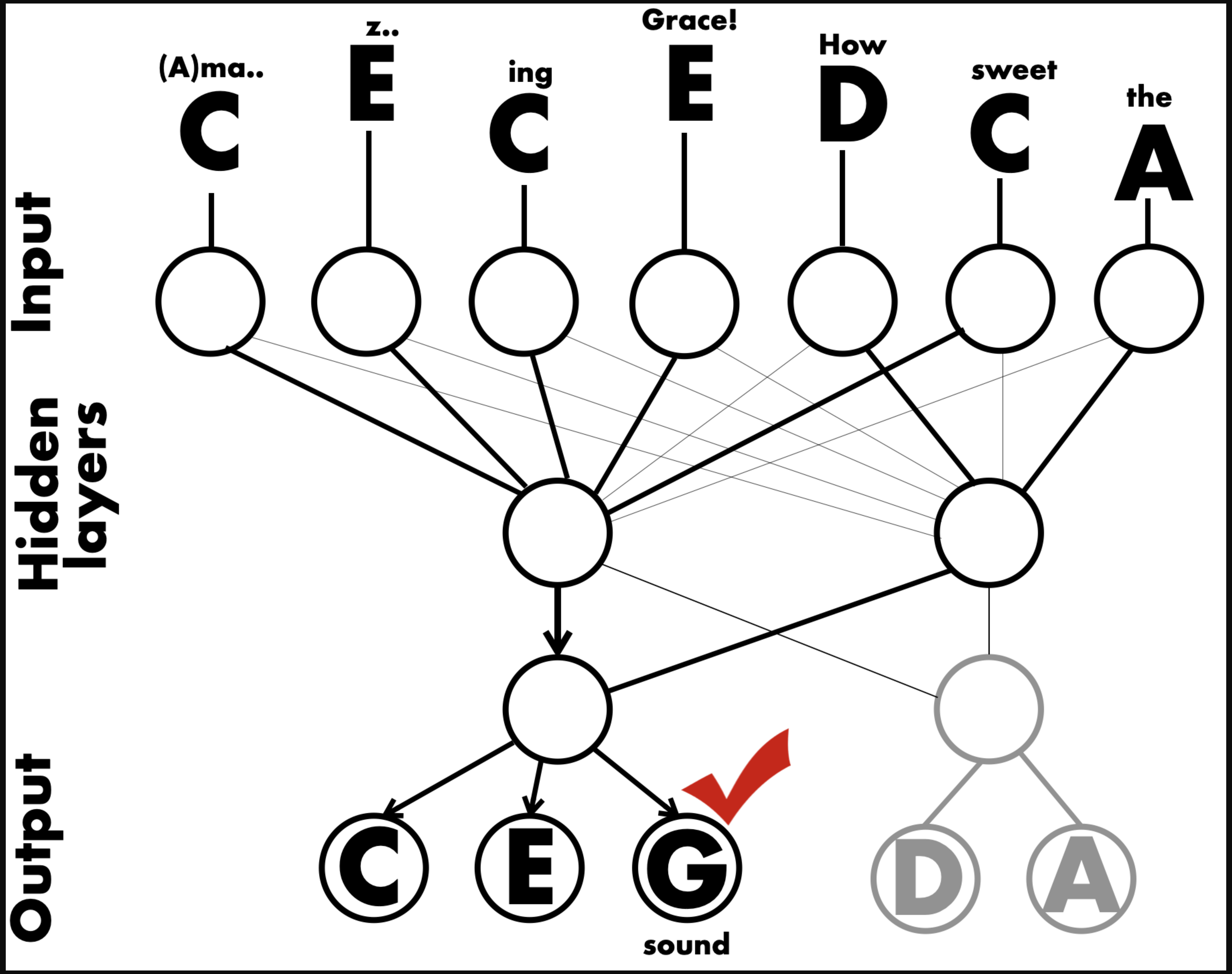
- A transformer is a type of neural network



First some definitions

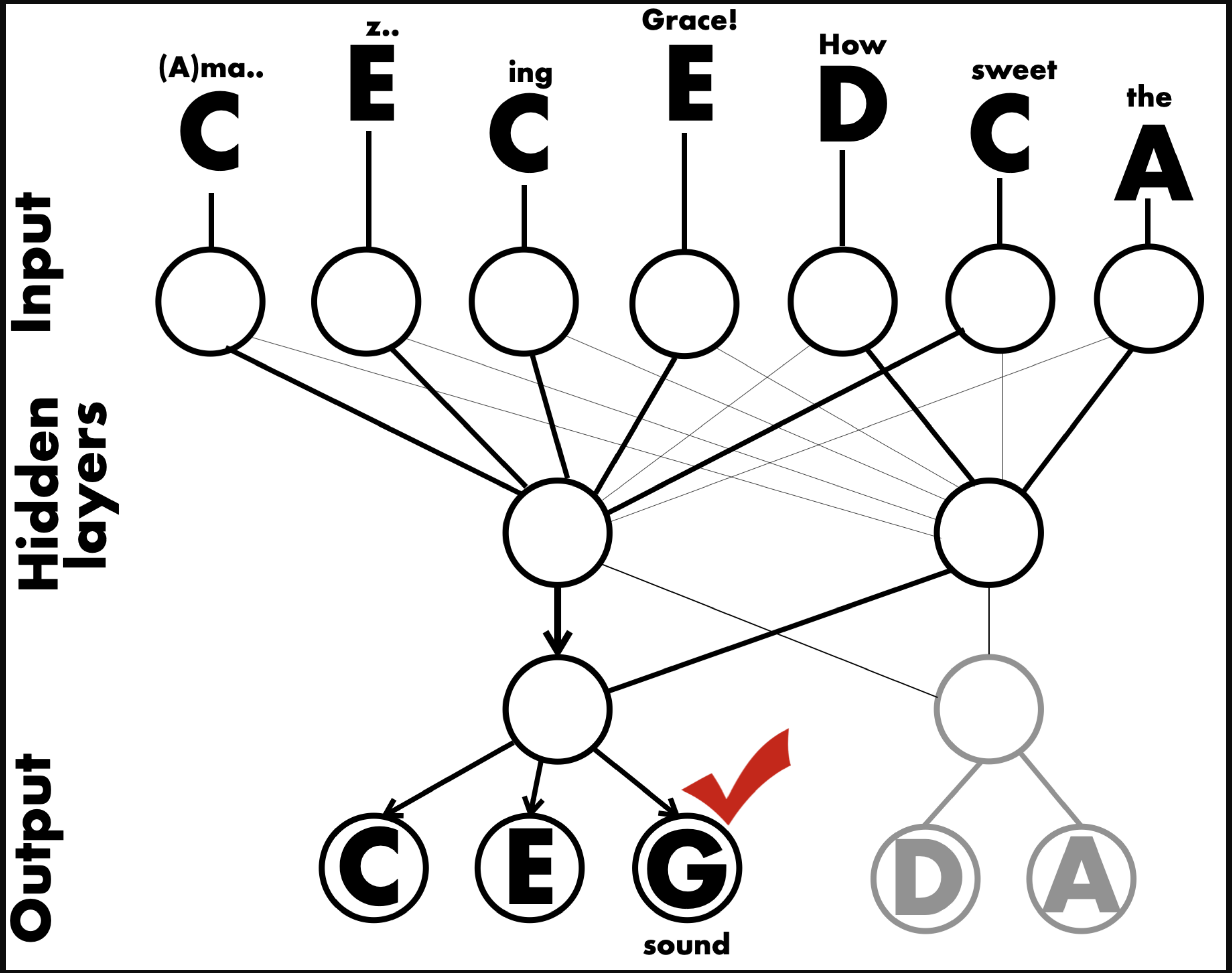
- Neural Network: A deep-learning system that uses multiple layers to capture complex connections between some observed input and output

- A transformer is a type of neural network



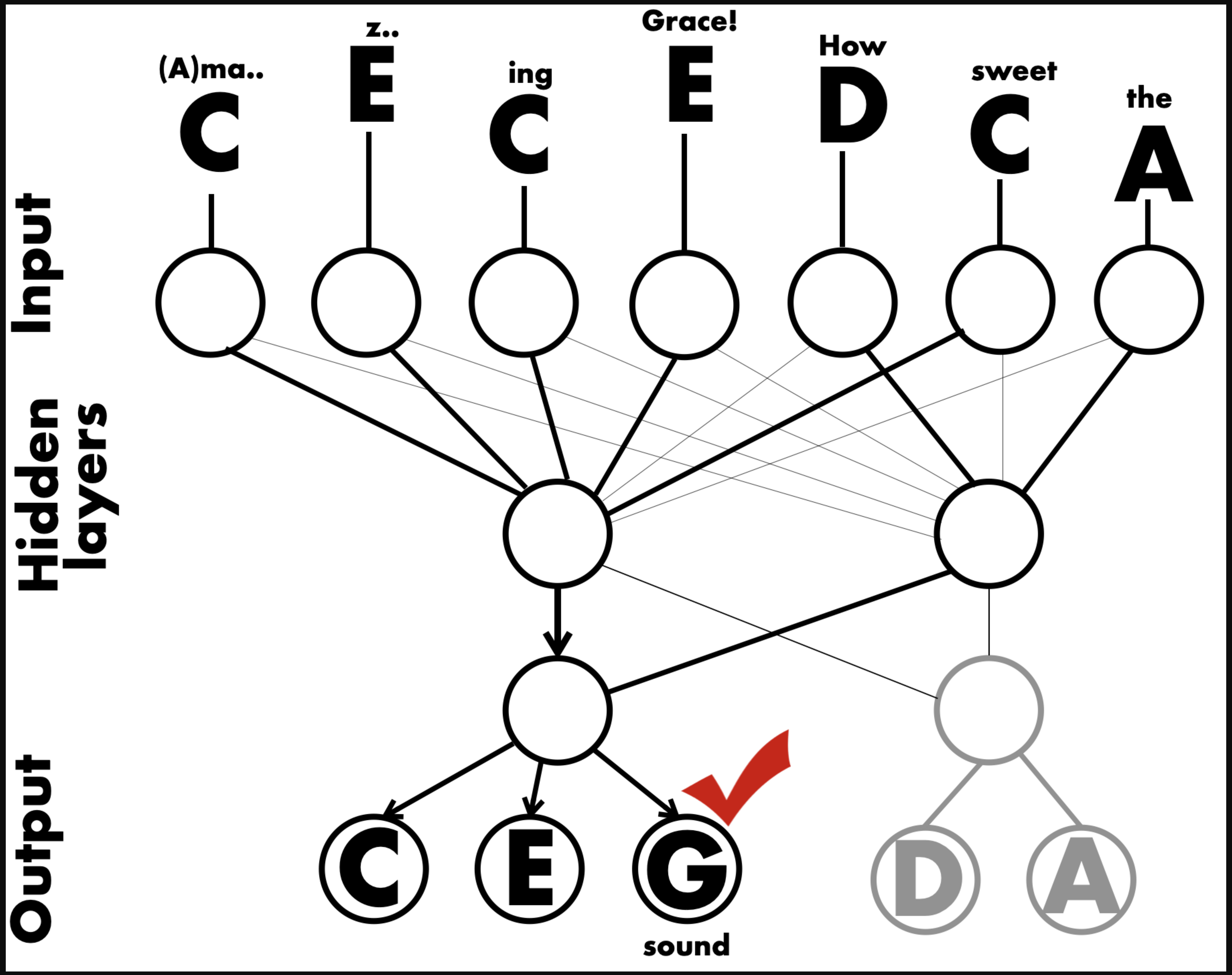
First some definitions

- To "train" (i.e., to machine-learn), it uses a dataset to set up lots of situations like this
  - If the Network guesses right, it strengthens that prediction
  - If if guesses wrong, it weakens that prediction
- After it has trained, it can create new content using these honed predictions

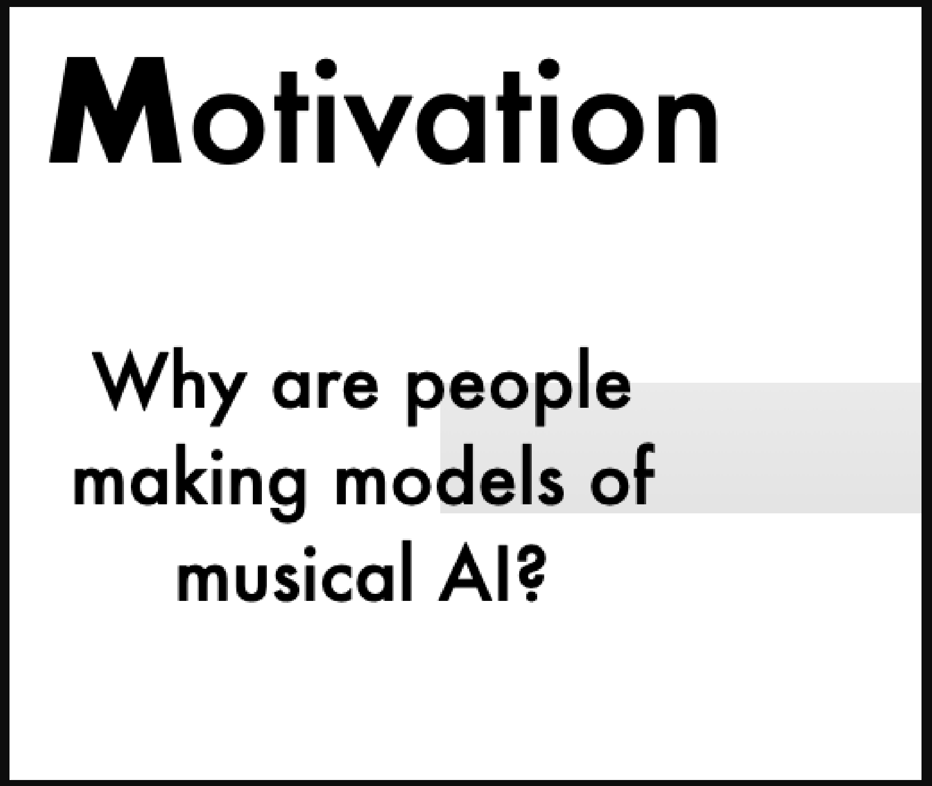


First some definitions

- Large Language Model: A neural network (usually a transformer) that has many layers and trains on a lot of data
- Generative AI is some machine-learned system that can create new content



Force #1: Motivations



Force #1: Motivations

Motivation

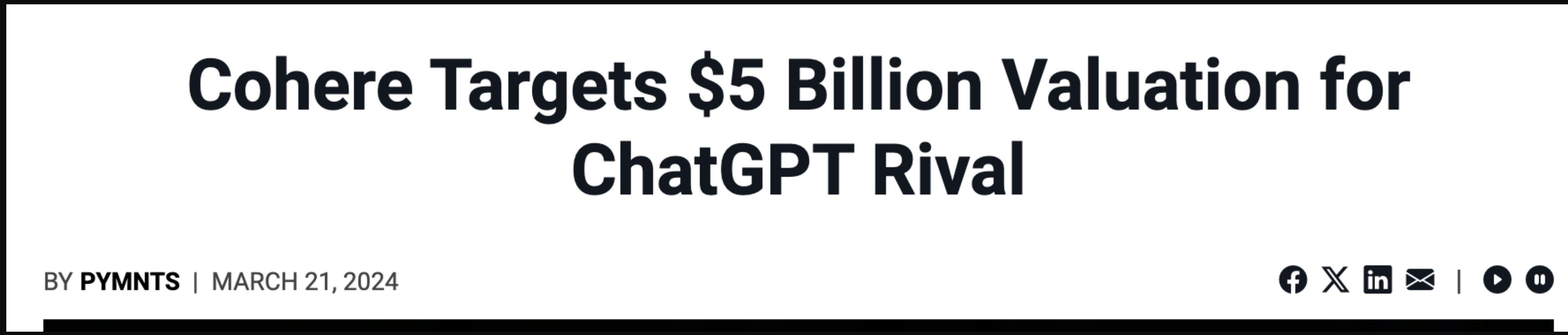
Why are people making models of musical AI?

Force #1: Motivations

There's not a lot of time and resources allocated to musical AI, likely because folks don't think it'll make a lot of money

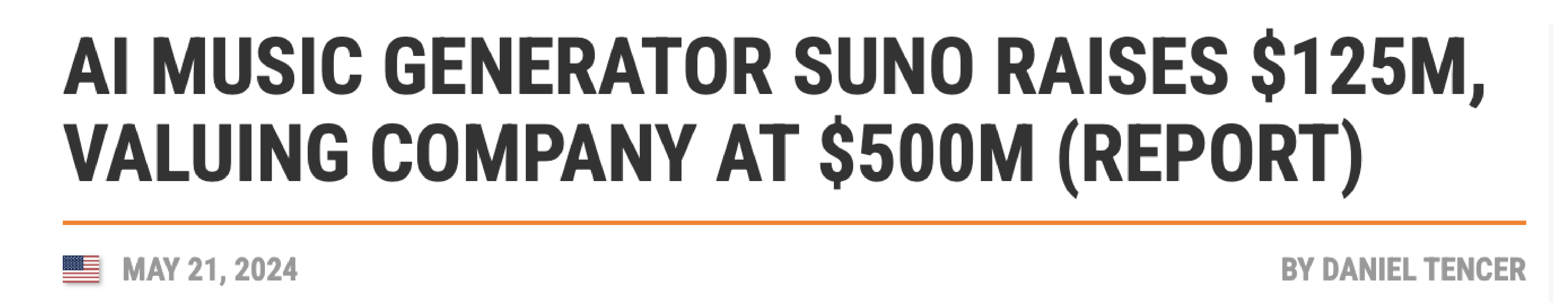
Motivations

- Within Generative AI, music is valued less on the market
- And within the overall market, Generative AI is valued less than other types of AI
- In other words, music gets the fewest resources


Cohere Targets $5 Billion Valuation for ChatGPT Rival

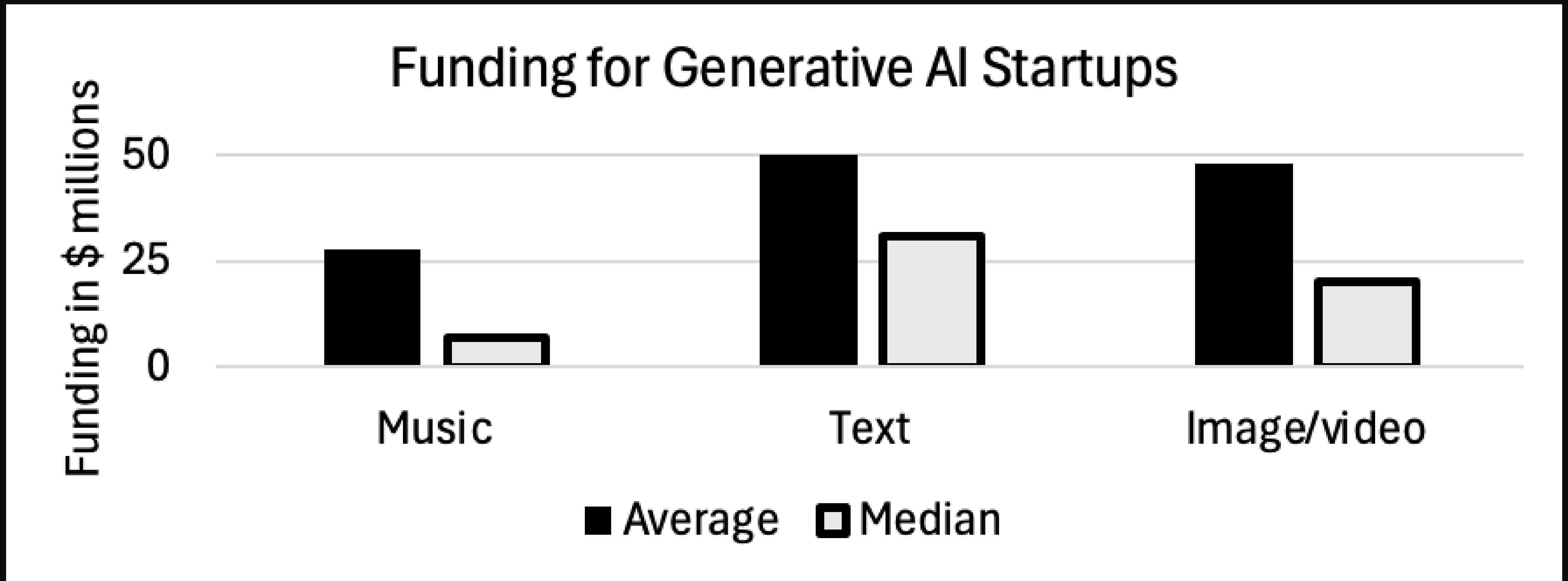
BY PYMNTS | MARCH 21, 2024

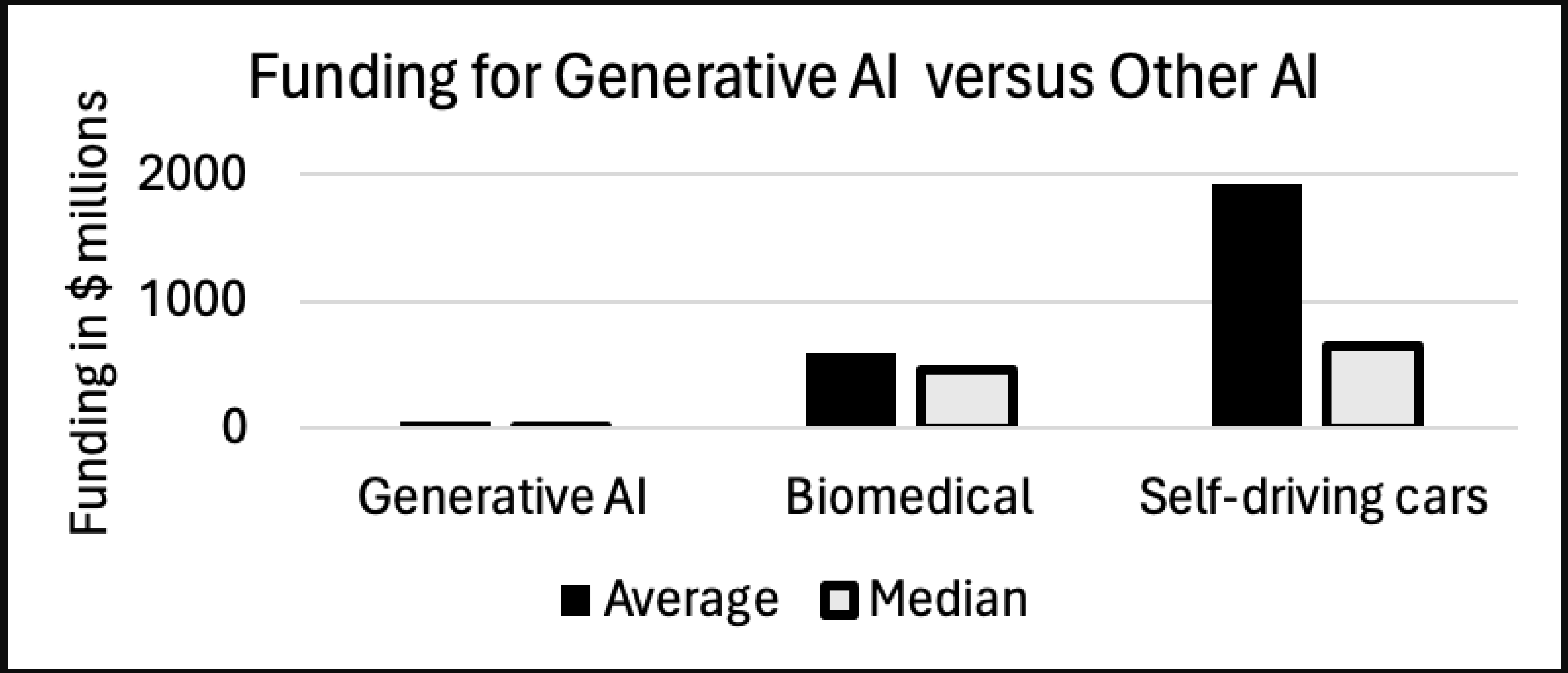
https://www.pymnts.com/artificial-intelligence-2/2024/cohere-targets-5-billion-valuation-for-chatgpt-rival/


AI MUSIC GENERATOR SUNO RAISES $125M, VALUING COMPANY AT $500M (REPORT)

MAY 21, 2024 BY DANIEL TENCER

https://www.musicbusinessworldwide.com/ai-music-generator-suno-raises-125m-valuing-company-at-500m-report/





Motivations

- And it costs so much to make and train one of these AIs, there needs to be some clear payoff
- And, given the lack of investment, it seems like market forces are skeptical that there’s as much money to be made in generative musical AI as other in domains of AI

MACHINES

The cost of training AI is surging, report warns

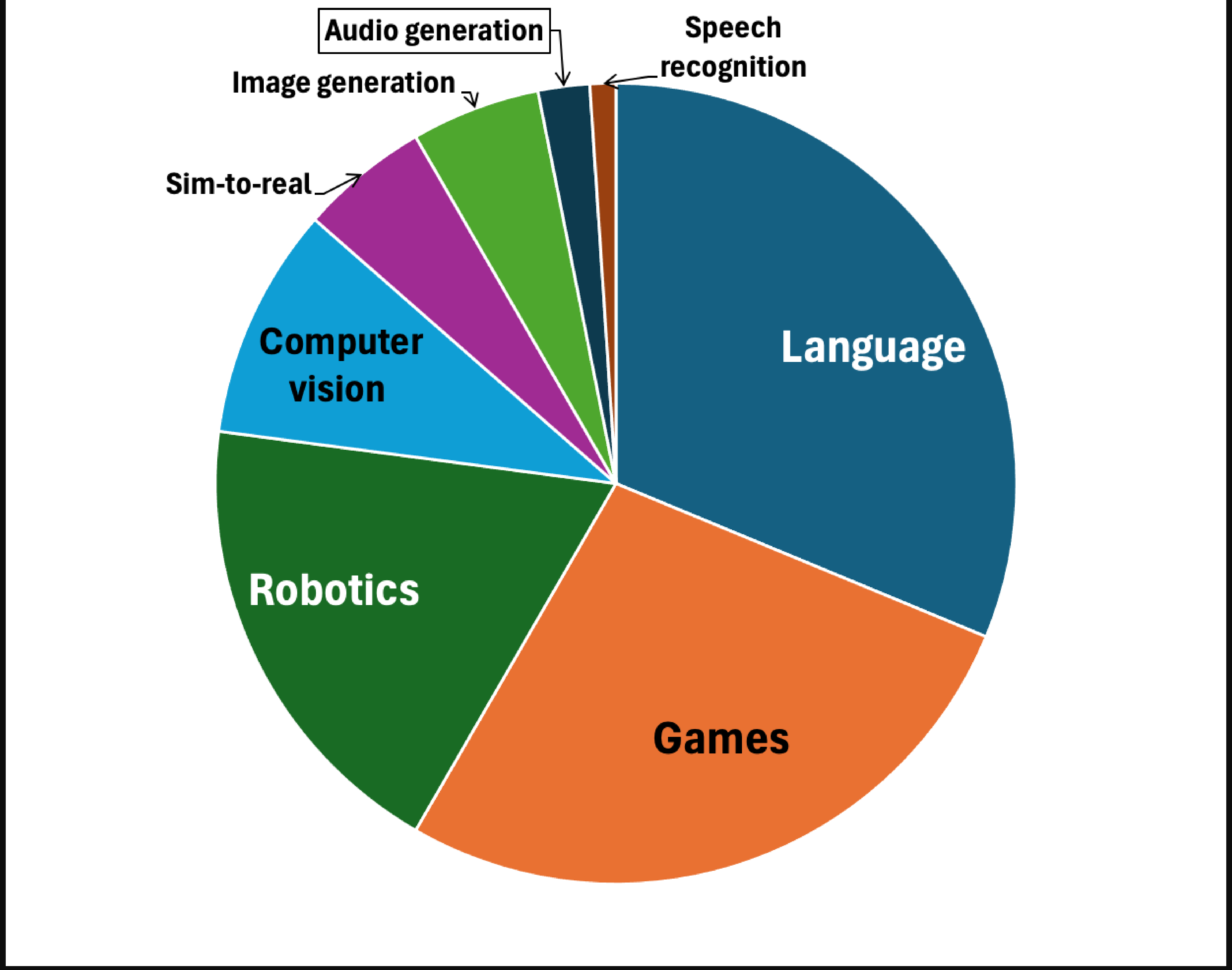
by Leigh Mc Gowran

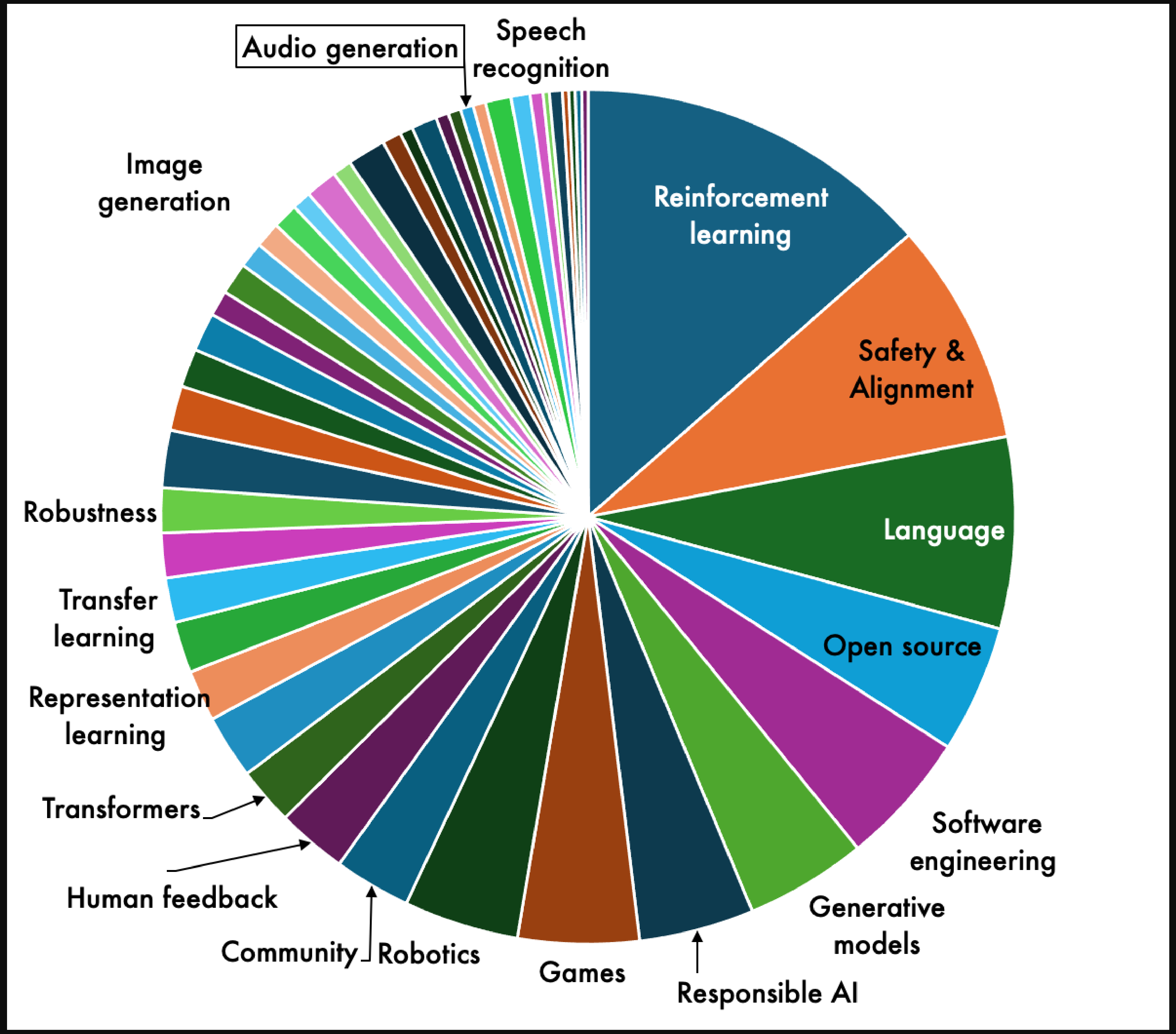
16 APR 2024 SAVE ARTICLE

Motivations

- Musical AI seems to be researched less as well

Papers published by OpenAI, a company that works on various different media:





Force #2: Examples

Motivation	Examples
Why are people making models of musical AI?	What datasets are people using to train and test their AI models?

Force #2: Examples

Musical datasets are smaller because it's hard to make music into data types that an AI can learn from

Examples:

LLMs need lots and lots of data to train on, and musical data is sparse and hard to acquire

Examples:

Problem #1: Musical datasets trend small

Examples:

Problem #1: Musical datasets trend small

Examples:

Problem #1: Musical datasets trend small

Examples:

Problem #1: Musical datasets trend small

Examples:

Problem #1: Musical datasets trend small

Examples:

Problem #1: Musical datasets trend small

Examples:

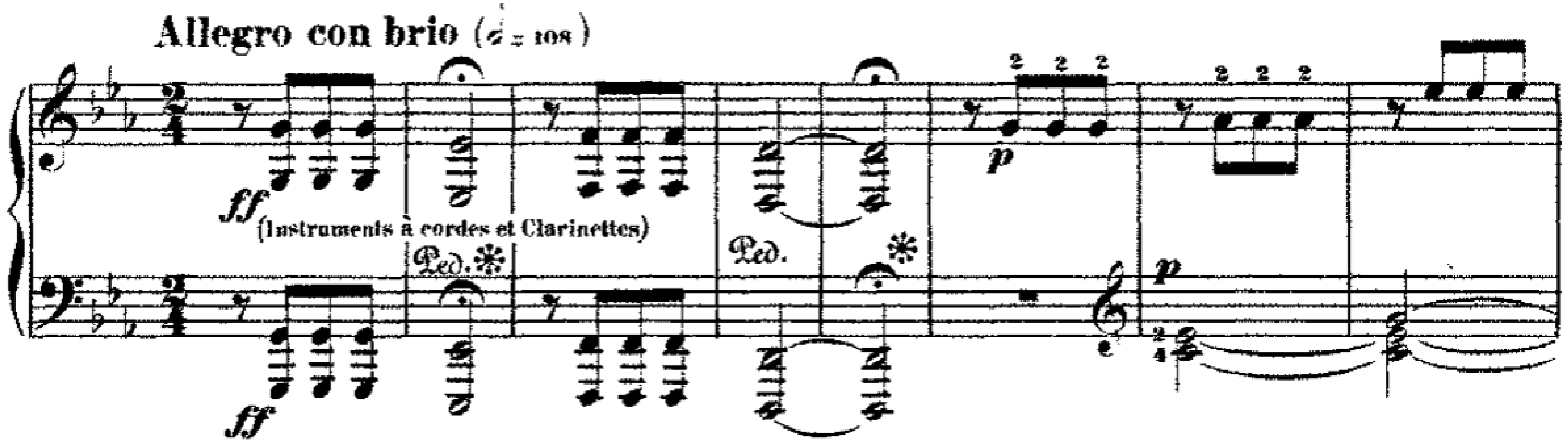
Problem #2: It's hard to extract computer readable information from pdfs and audio

Examples:

Problem #2: It's hard to extract computer readable information from pdfs and audio

The technology to extract reliable information from score images is way behind text recognition

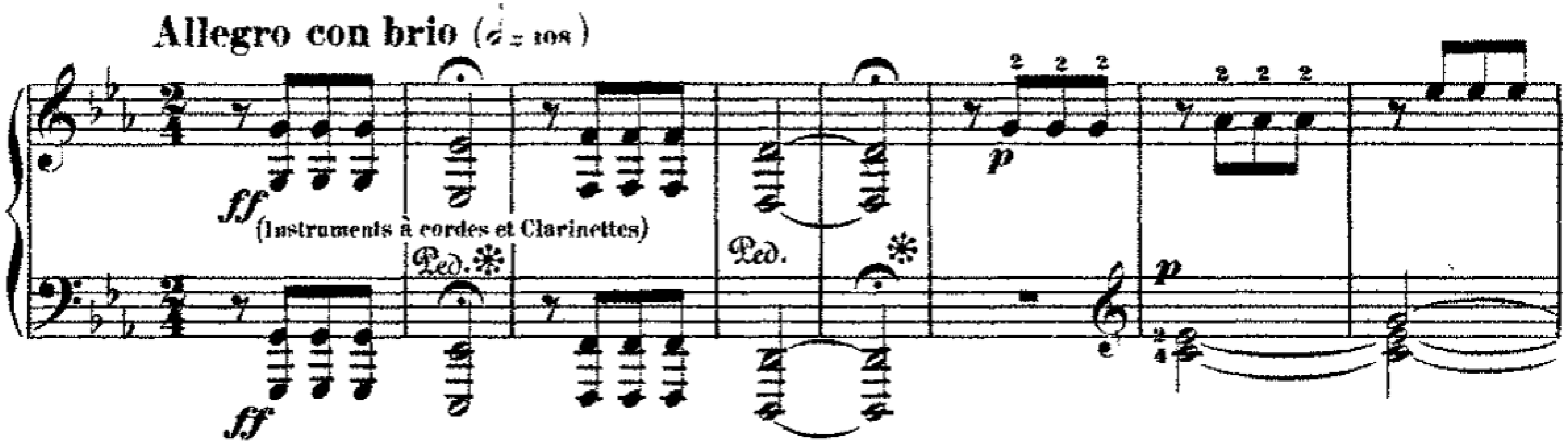
Original



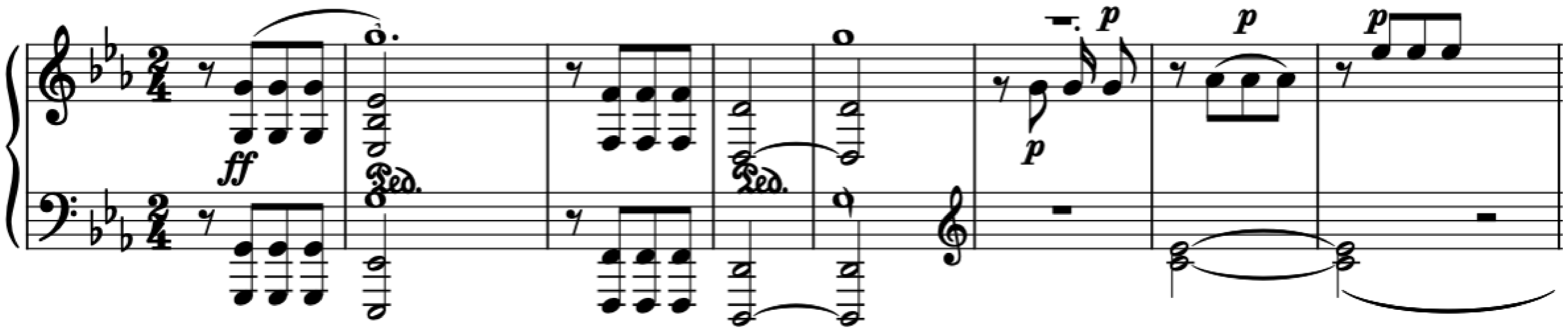
Examples: Problem #2 a

Examples: Problem #2 b

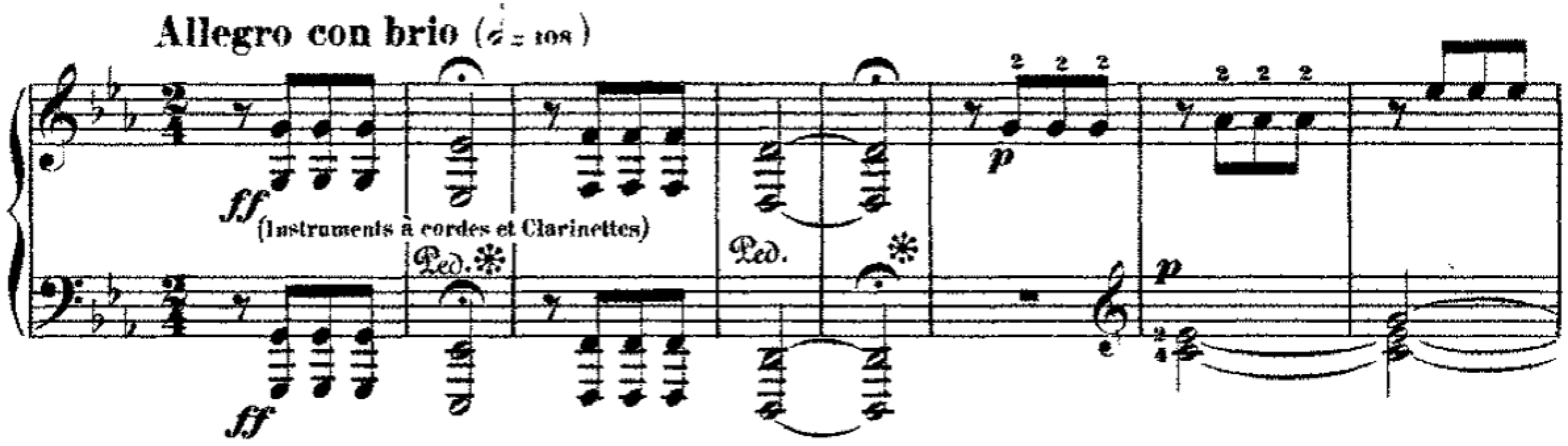
Original



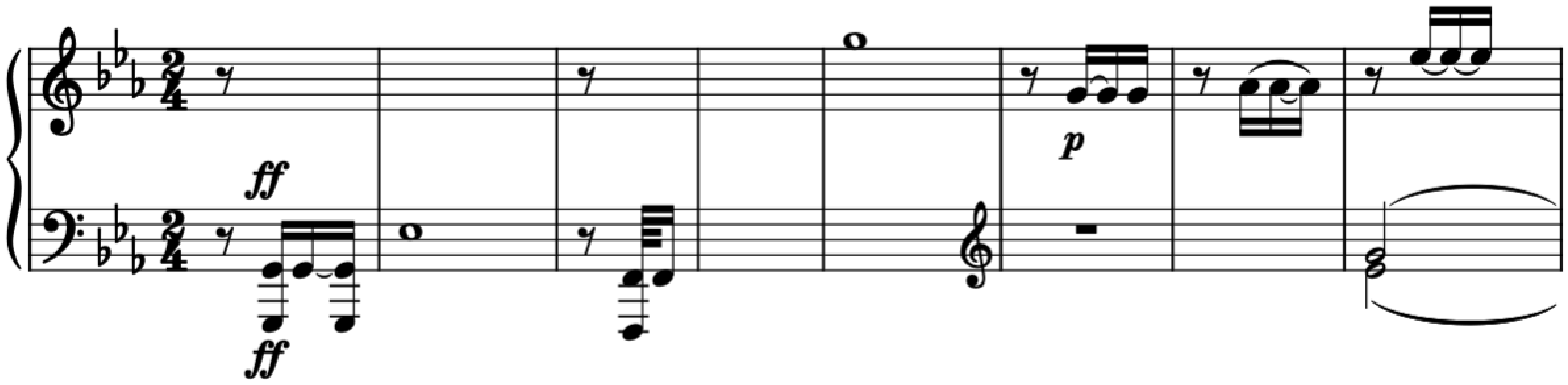
Scan 1: *includes several added notes*



Original

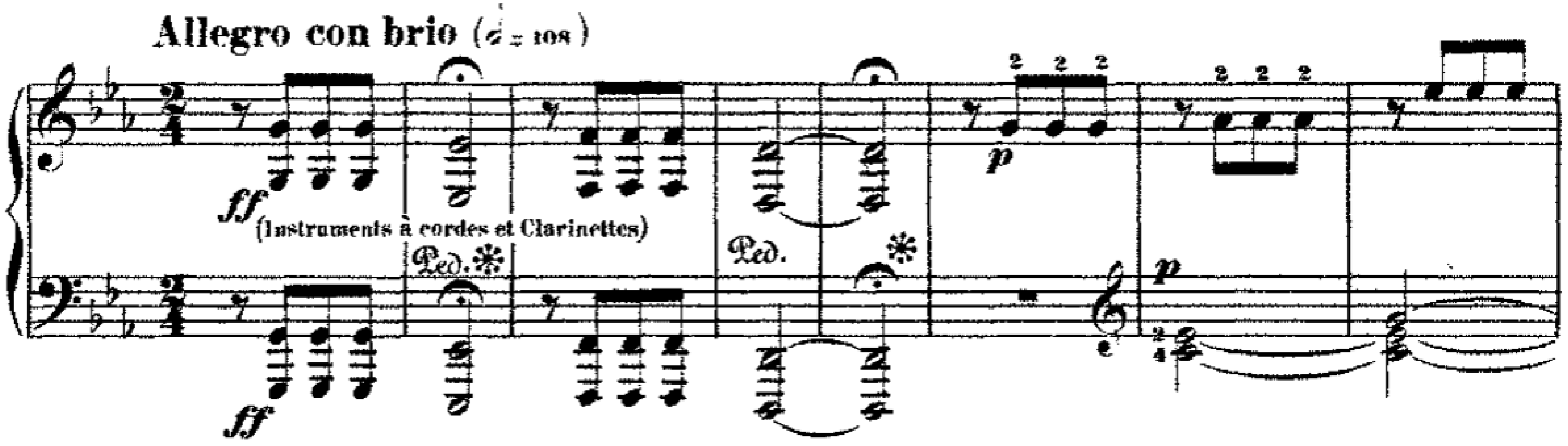


Scan 2: *includes several deleted noes*



Exan

Original



Scan 3:
links all staves on a page together

... etc (there are 5 more concurrent staves)

Yuck.

So let's use audio!

How about…

Lizzo's "Truth Hurts"

Hold on, let's talk to 20 seconds about overtones

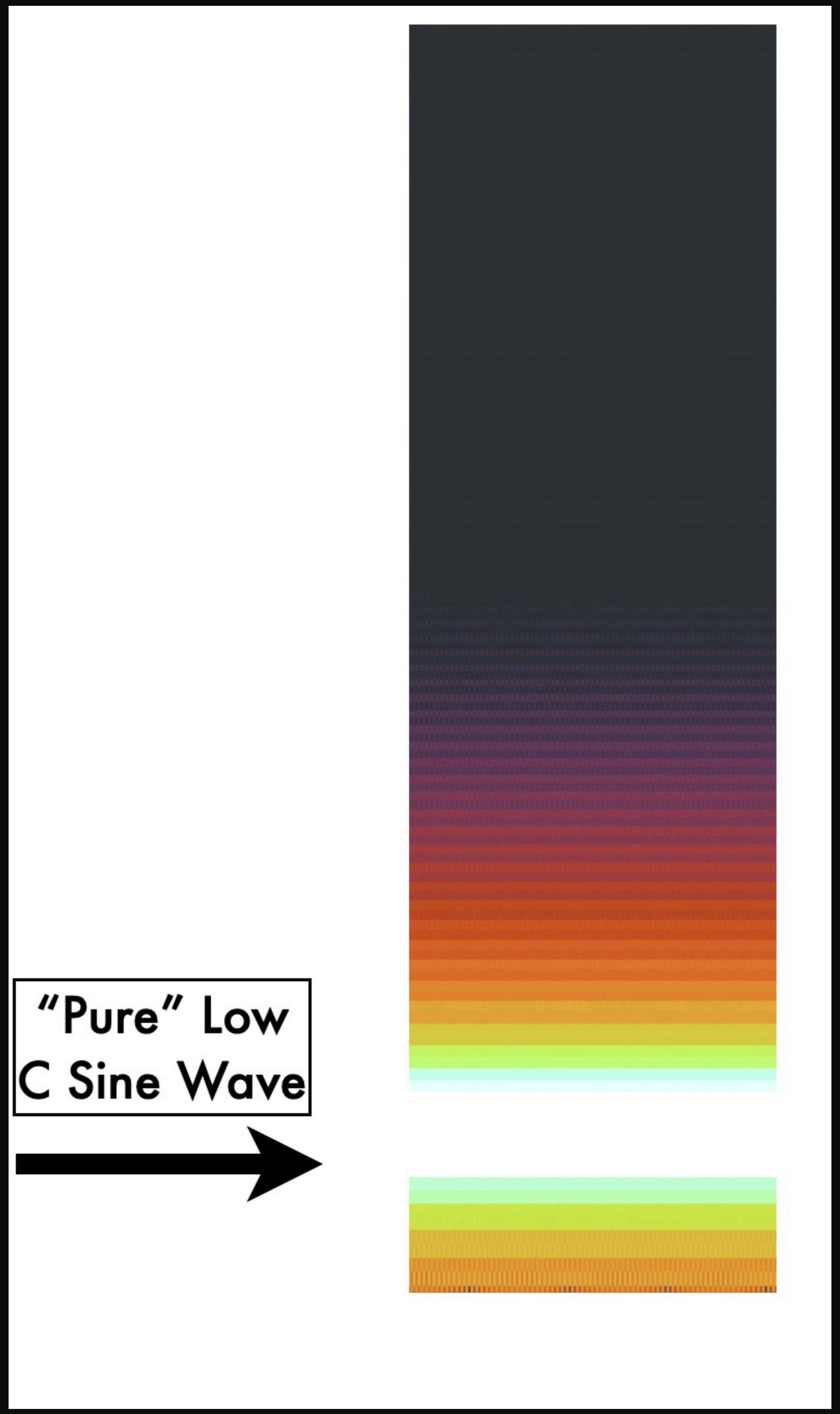
Complex and rich sounds contain complentary sound waves that create their unique timbre and quality

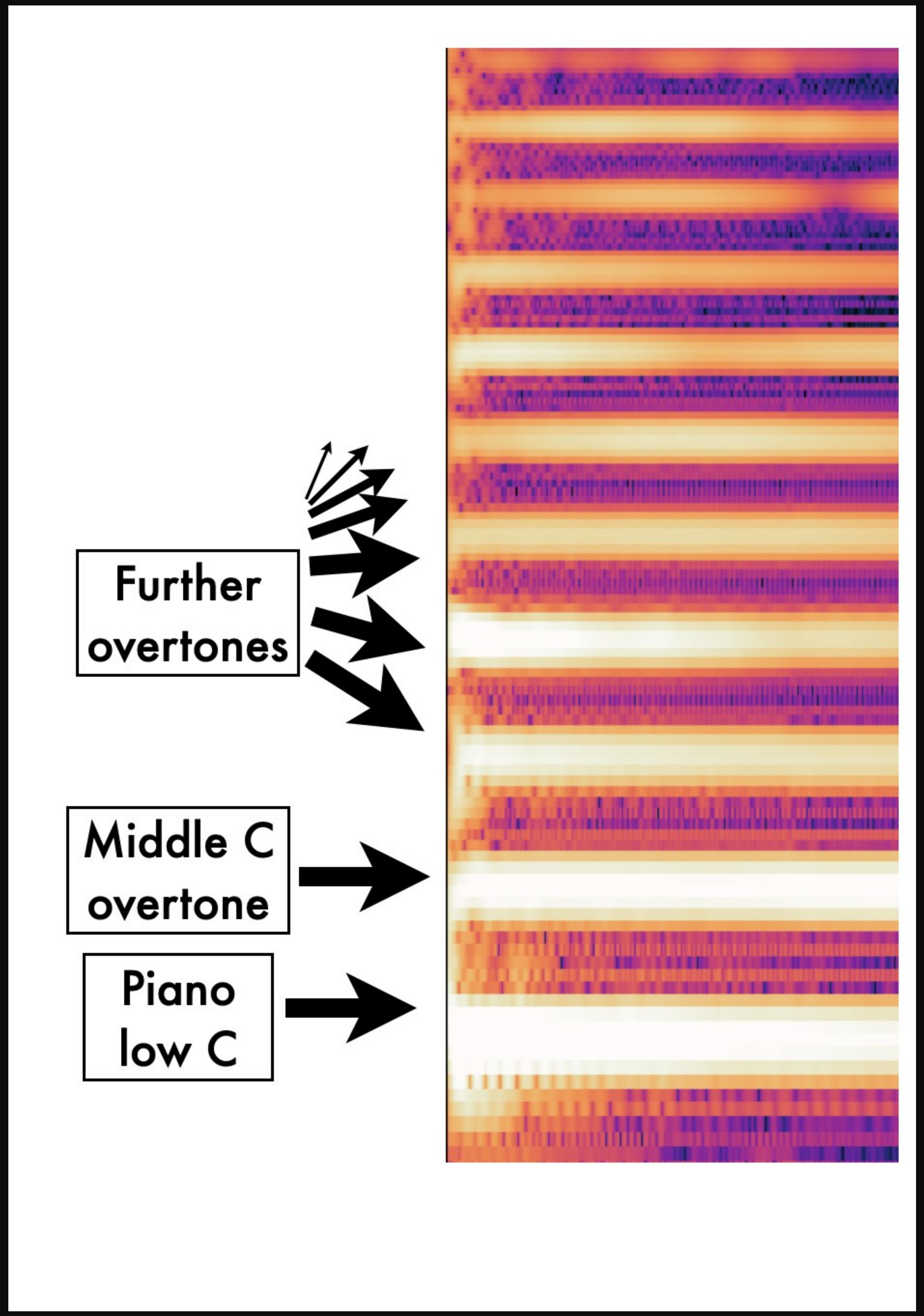
Overtones a

The horizontal axis is time, the vertical axis is frequency…

…so higher pitches are higher on the page

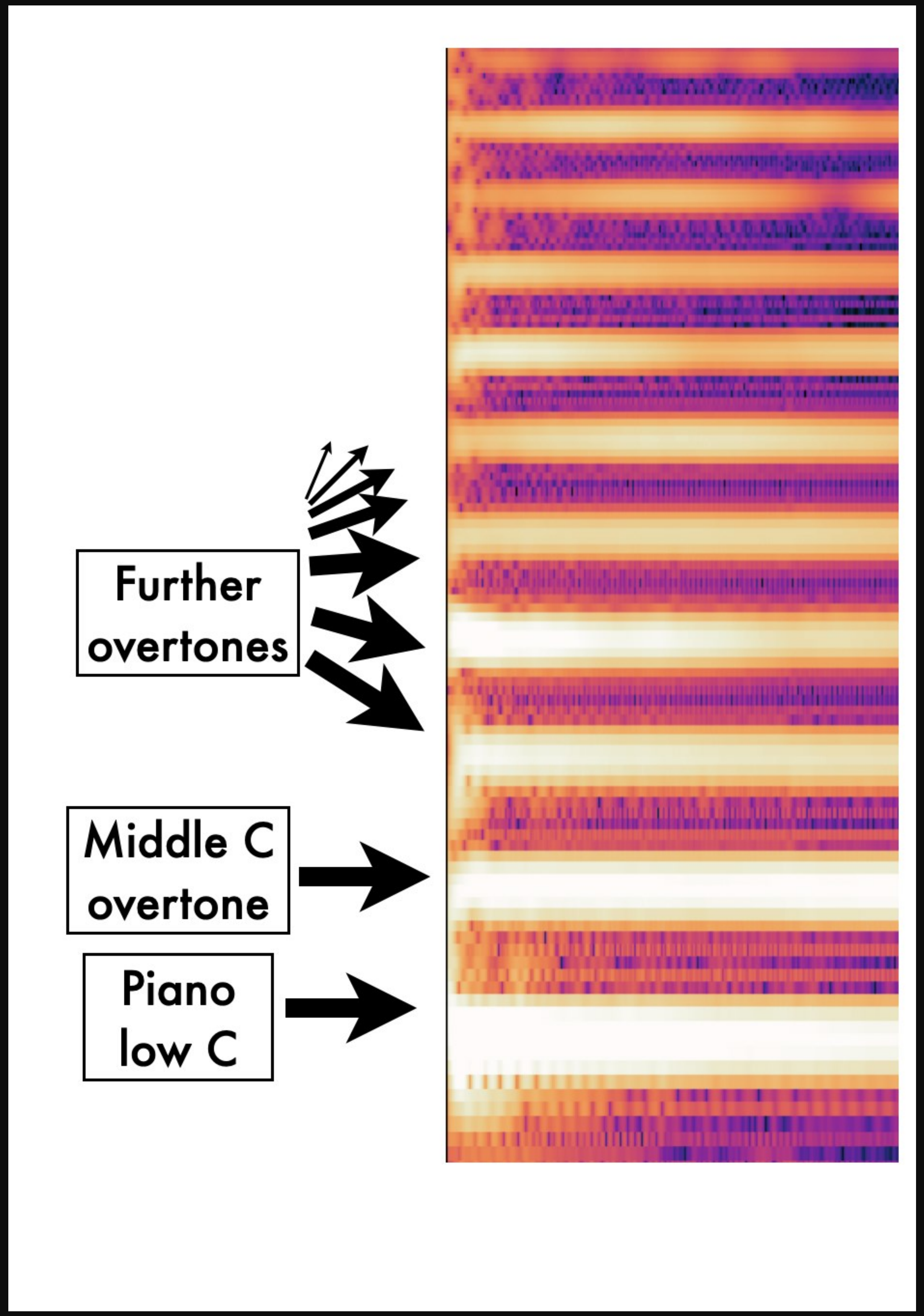
This lets you visualize all the complex overtones and complementary sounds that go into a rich timbre





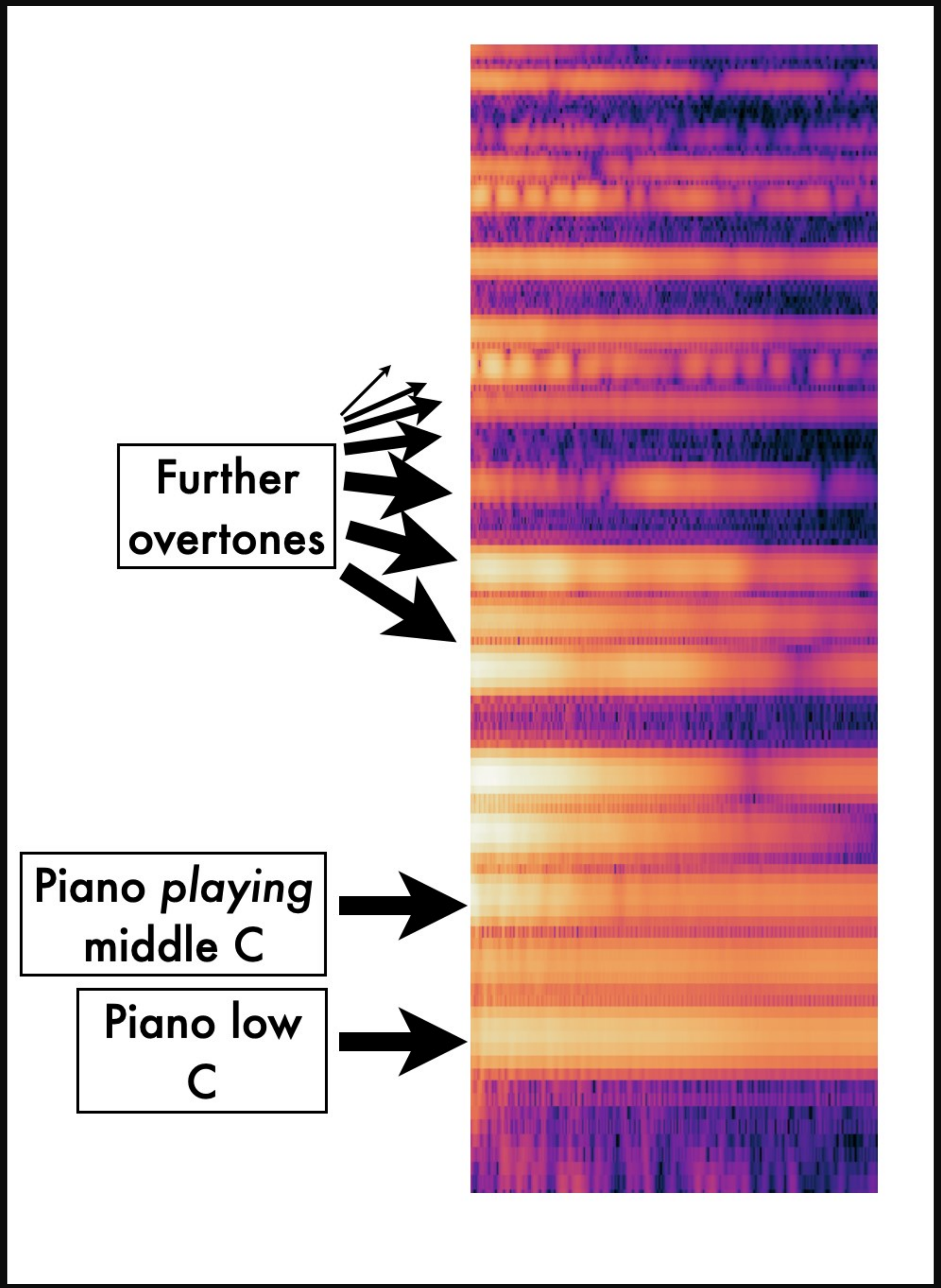
Overtones b

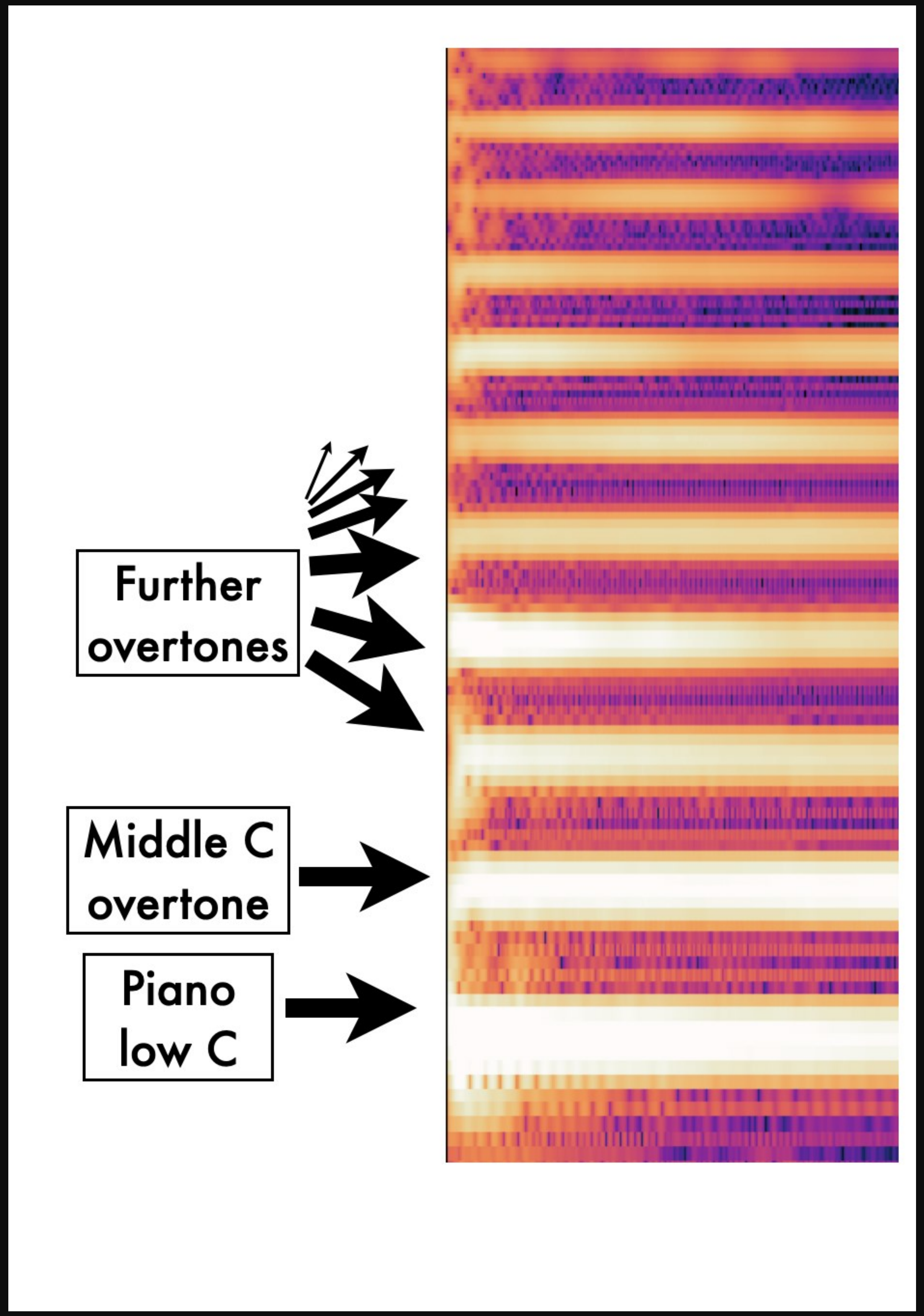
Here’s the problem: overtones look a lot like other notes



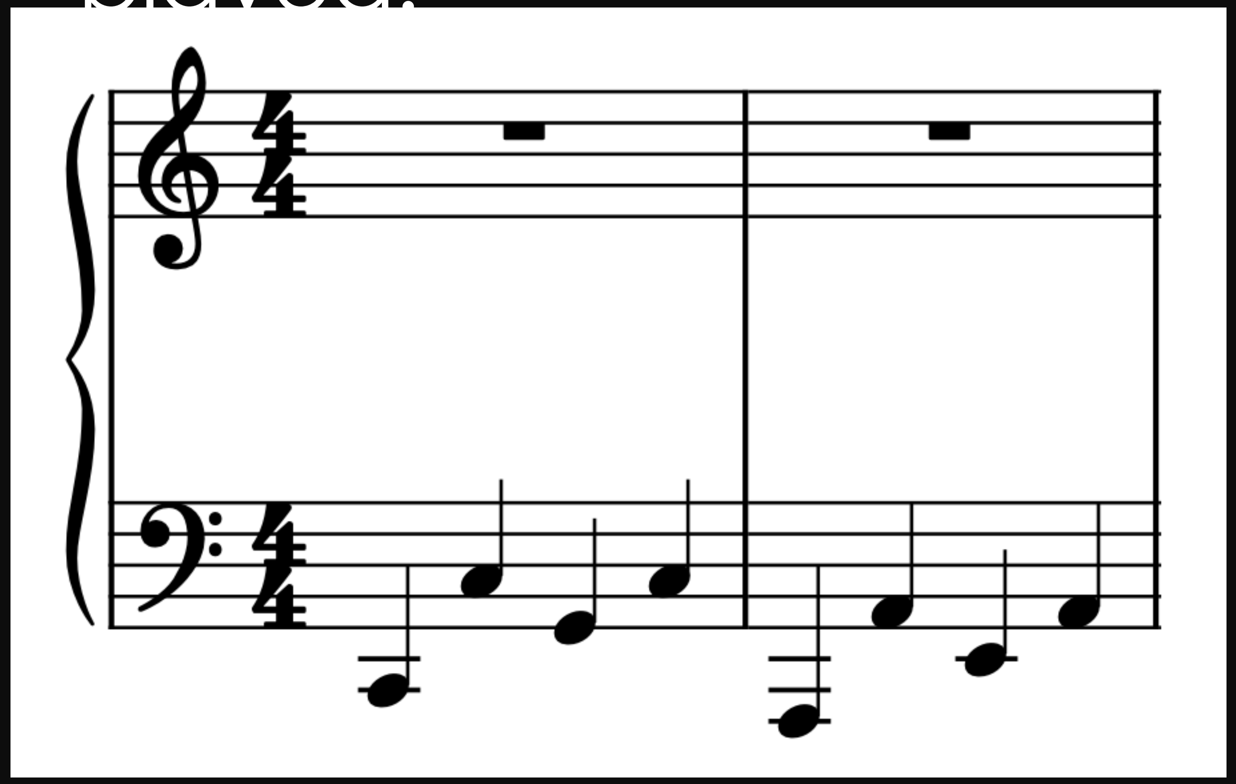
Overtones c

Here's the problem: overtones look a lot like other notes

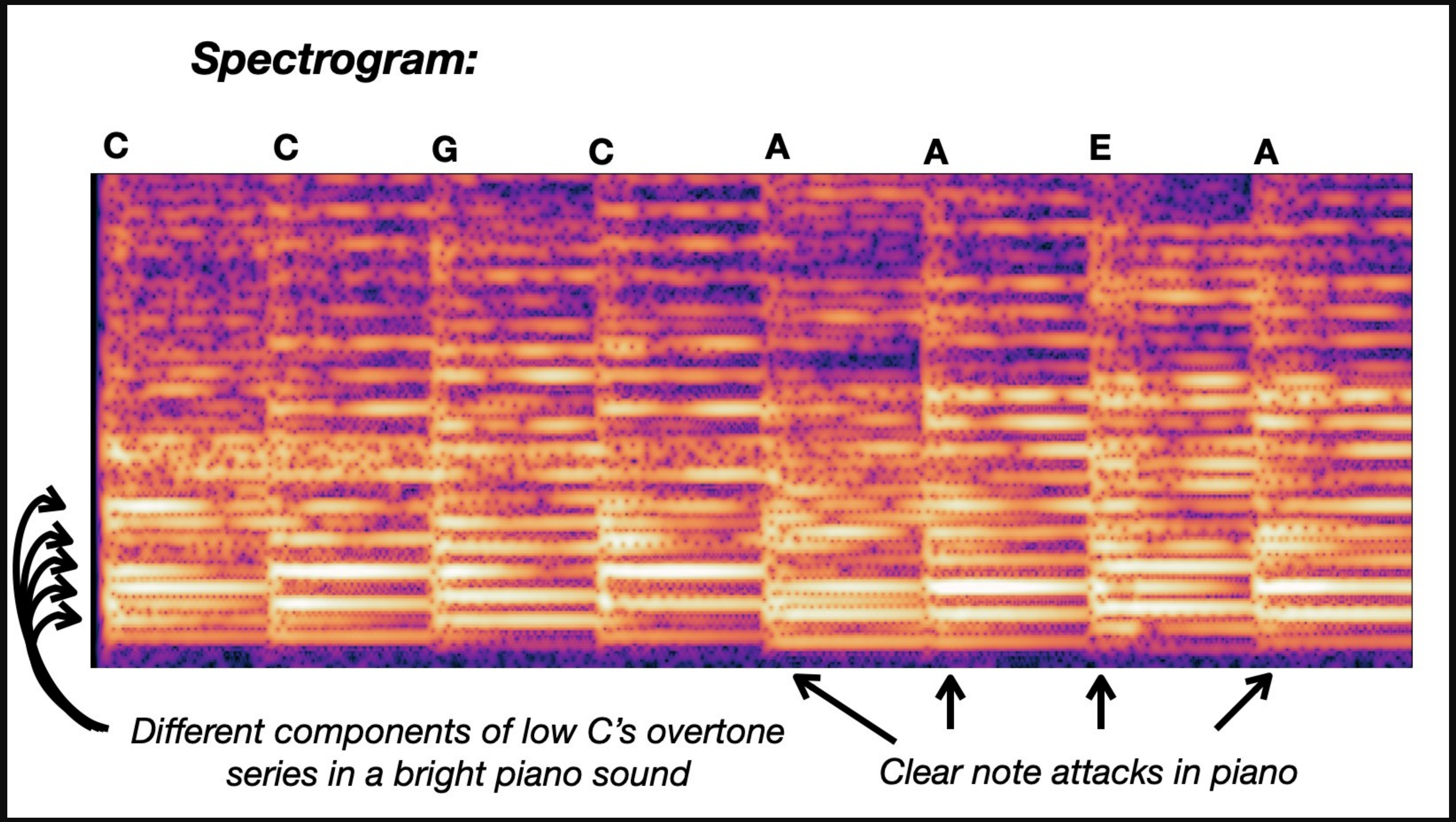




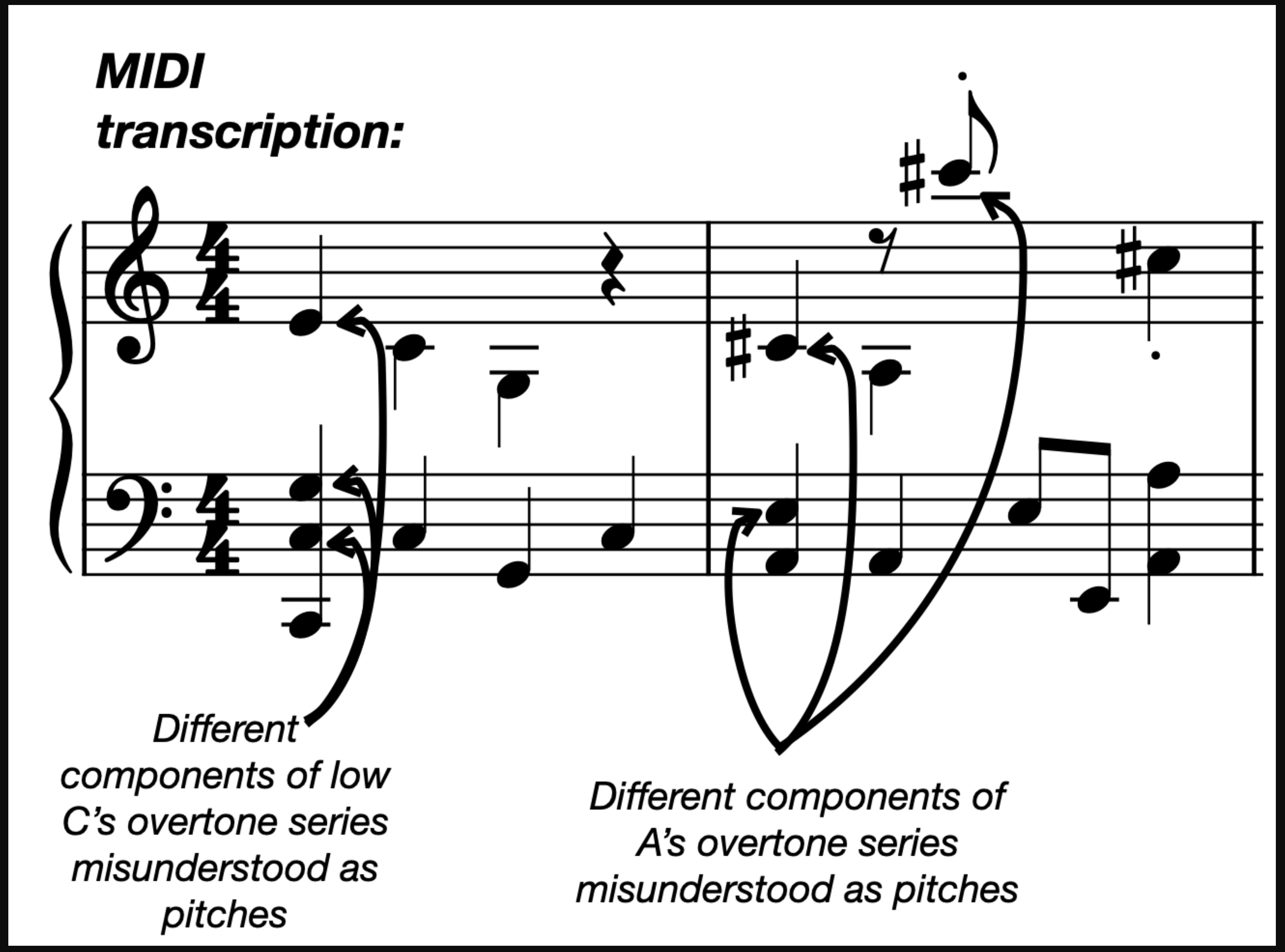
What's actually played:



The audio signal looks like:



The transcription looks like:



Force #3: Representation

Motivation	Examples	Representation
Why are people making models of musical AI?	What datasets are people using to train and test their AI models?	How are programmers representing musical events in their AI?

Force #3: Representation

Motivation	Examples	Representation
Why are people making models of musical AI?	What datasets are people using to train and test their AI models?	How are programmers representing musical events in their AI?

Force #3: Representation

Music has so many interrelated moving parts that it's hard to cleanly chunk up musical events in ways that an AI can learn from and use

Representation:

You can't just turn on the radio and yell at your LLM to start listening

You need to chunk the data up into tokens that the deep learning process can analyze and learn from

However you tokenize your data is how the medium will be represented in the AI's "mind"

Representations

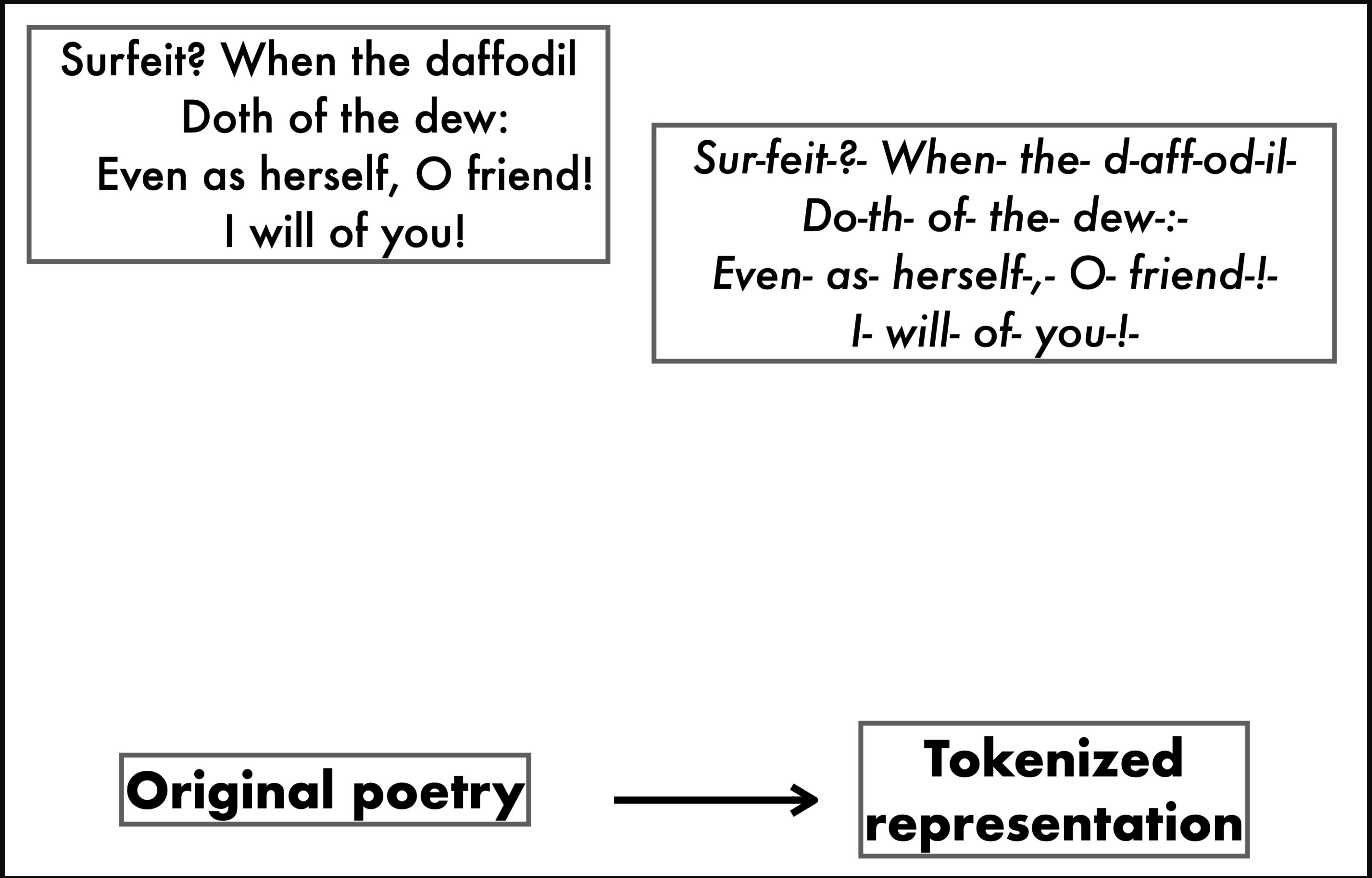
For instance, ChatGPT tokenizes using word chunks. This Emily Dickinson poem within its dataset would be tokenized like so:

Representations

For instance, ChatGPT tokenizes using word chunks. This Emily Dickinson poem within its dataset would be tokenized like so:

Representations

For instance, ChatGPT tokenizes using word chunks. This Emily Dickinson poem within its dataset would be tokenized like so:

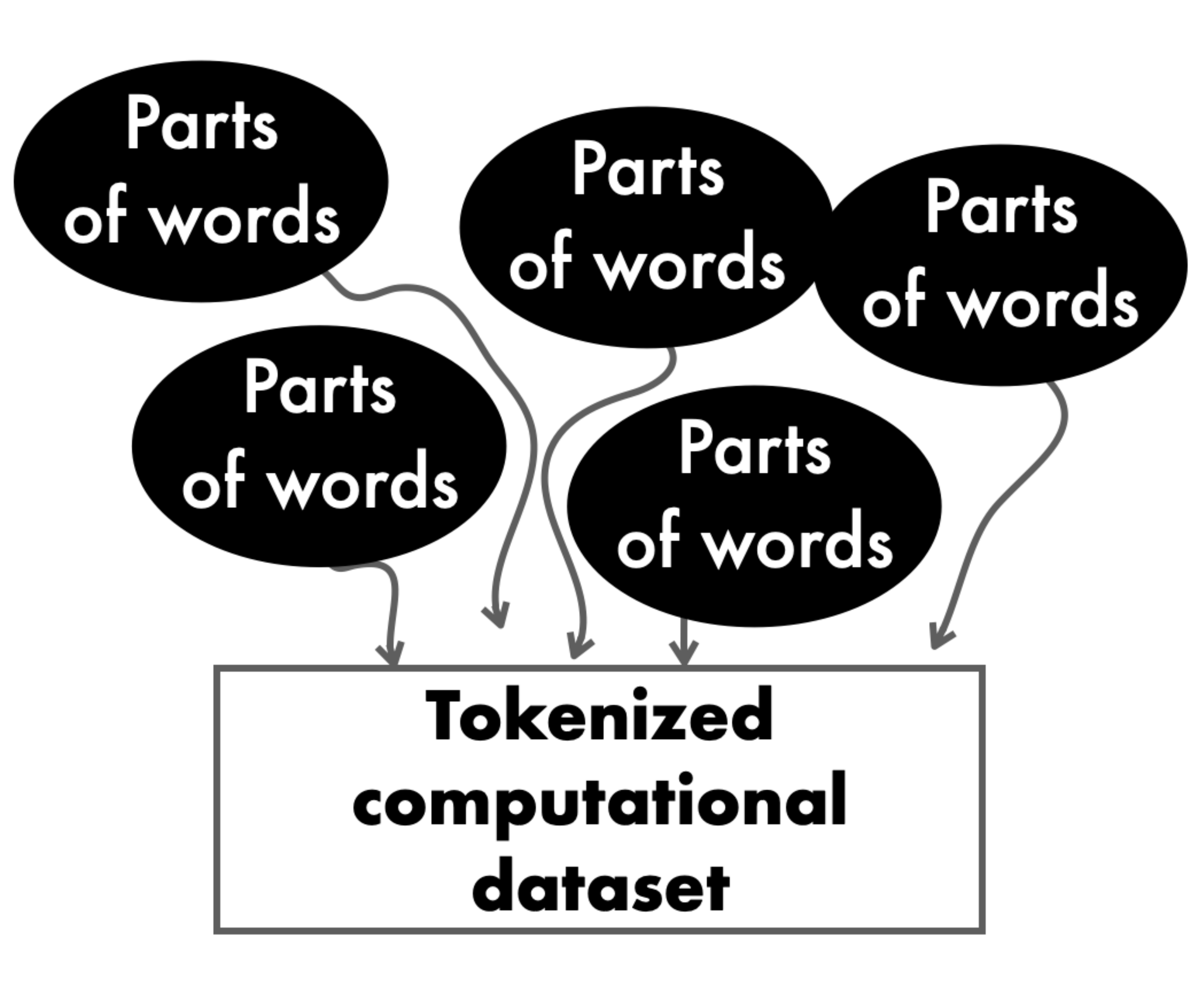


Representations

ChatGPT's mind (its neural network) knows then how these word chunks should be strung together, and how to arrange them into patterns similar to other poetry it's seen

Representations

ChatGPT's mind (its neural network) knows then how these word chunks should be strung together, and how to arrange them into patterns similar to other poetry it's seen



Representations

ChatGPT then produces new poems by stringing these word chunks into new patterns that get stitched together for the user:

Representations

ChatGPT then produces new poems by stringing these word chunks into new patterns that get stitched together for the user:

“Make me a new poem in the style of Emily Dickinson!”

Representations

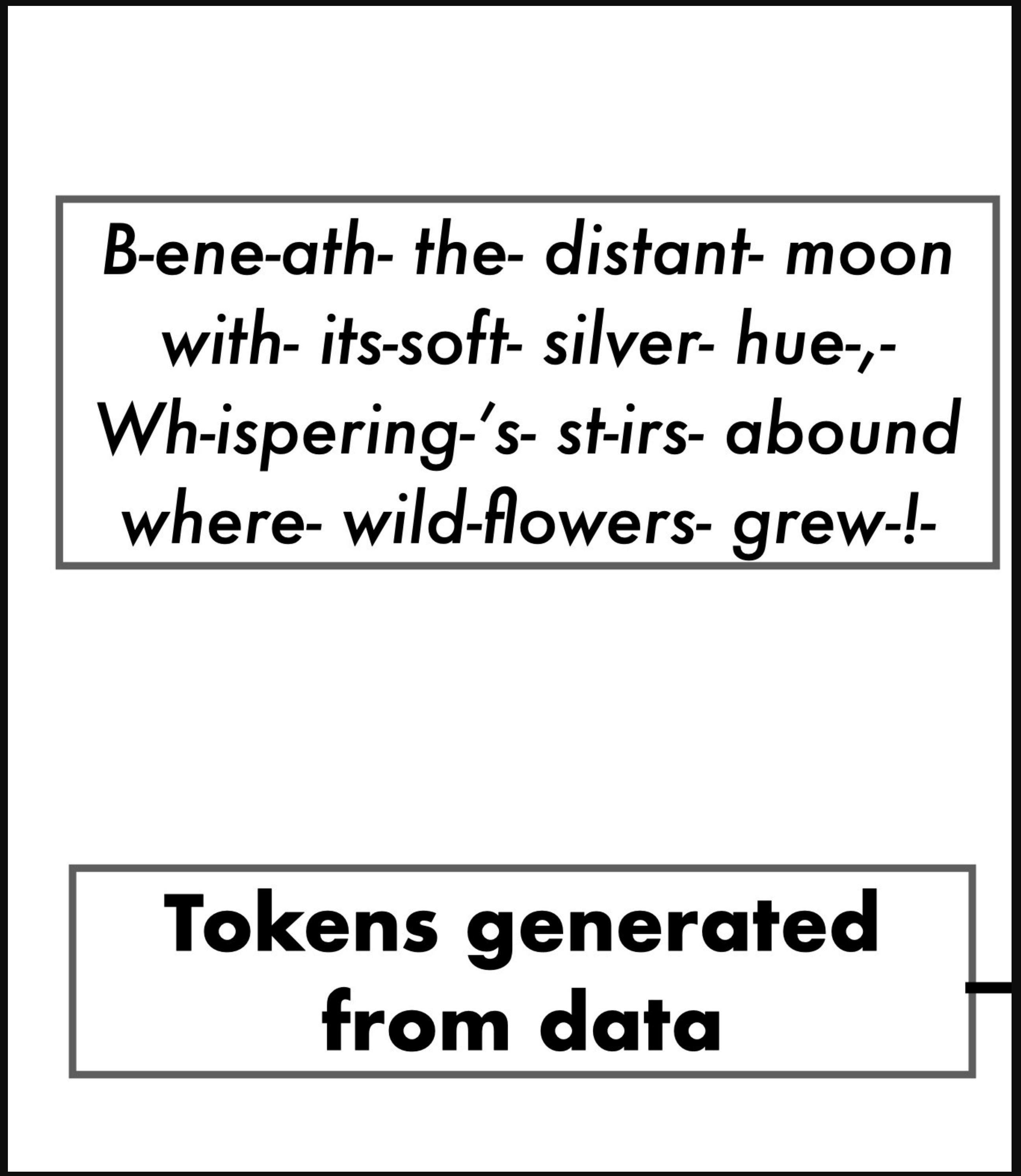
ChatGPT then produces new poems by stringing these word chunks into new patterns that get stitched together for the user:

"Make me a new poem in the style of Emily Dickinson!"

(Please)

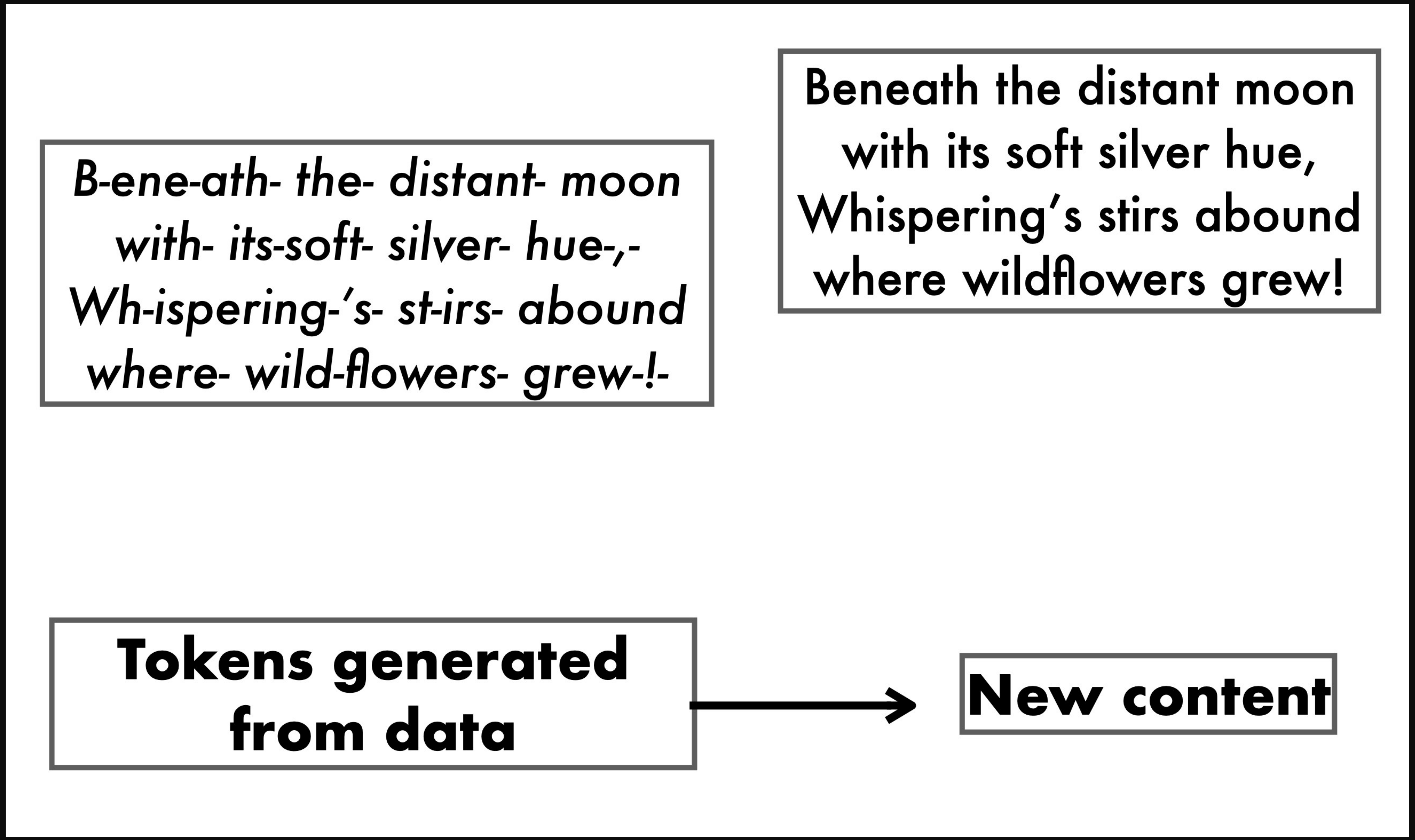
Representations

ChatGPT then produces new poems by stringing these word chunks into new patterns that get stitched together for the user:



Representations

ChatGPT then produces new poems by stringing these word chunks into new patterns that get stitched together for the user:

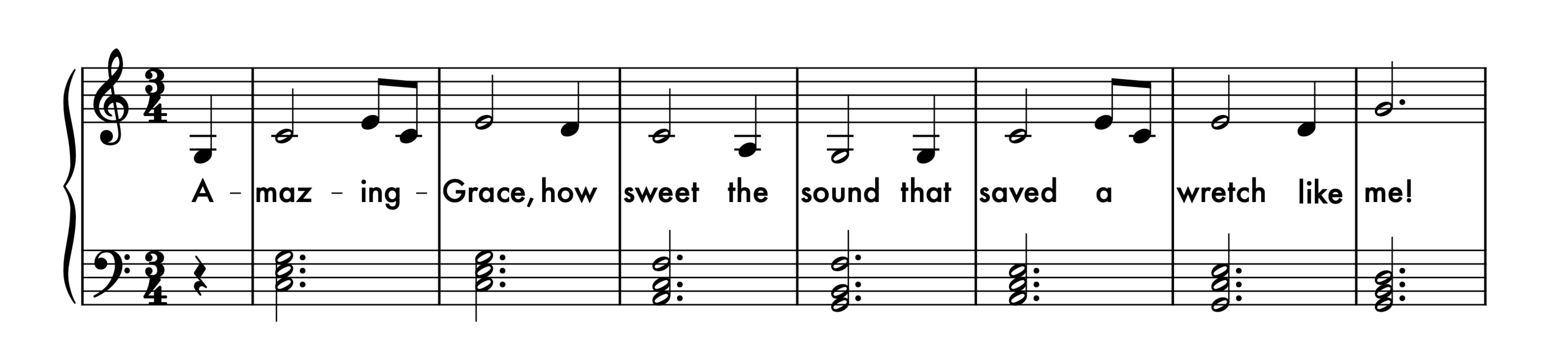


Representations

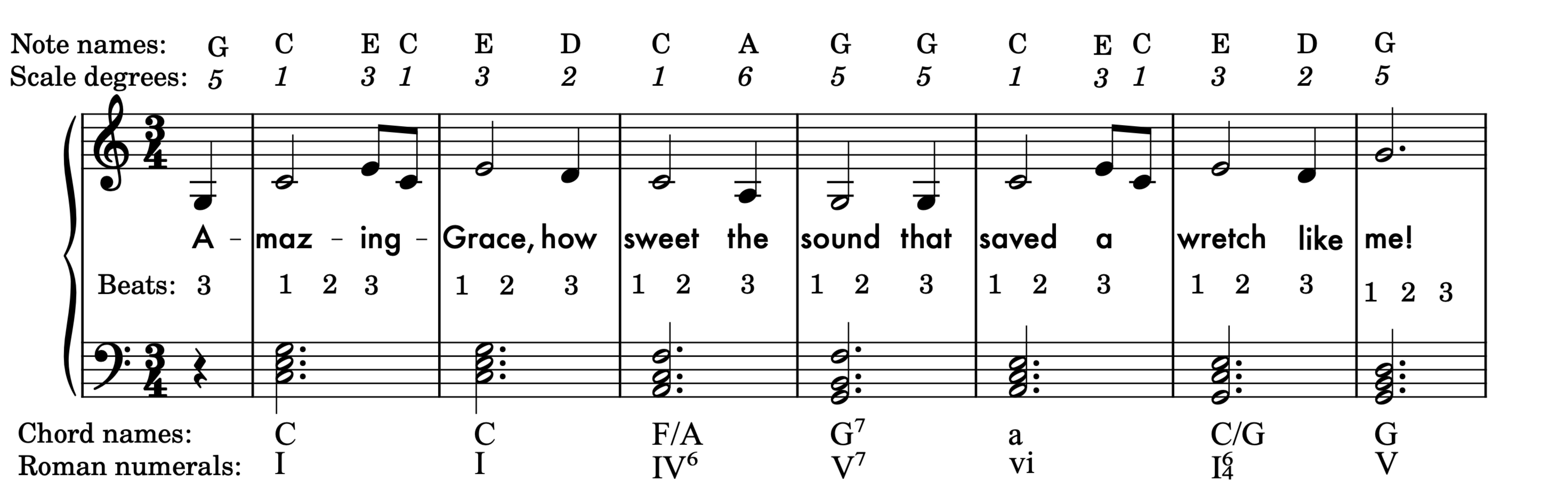
But, if you wanted to play around with other ways to tokenize text, your options would be obvious— letters, full words, phrases, or different chunks.

Musical tokenization is…. much less obvious.

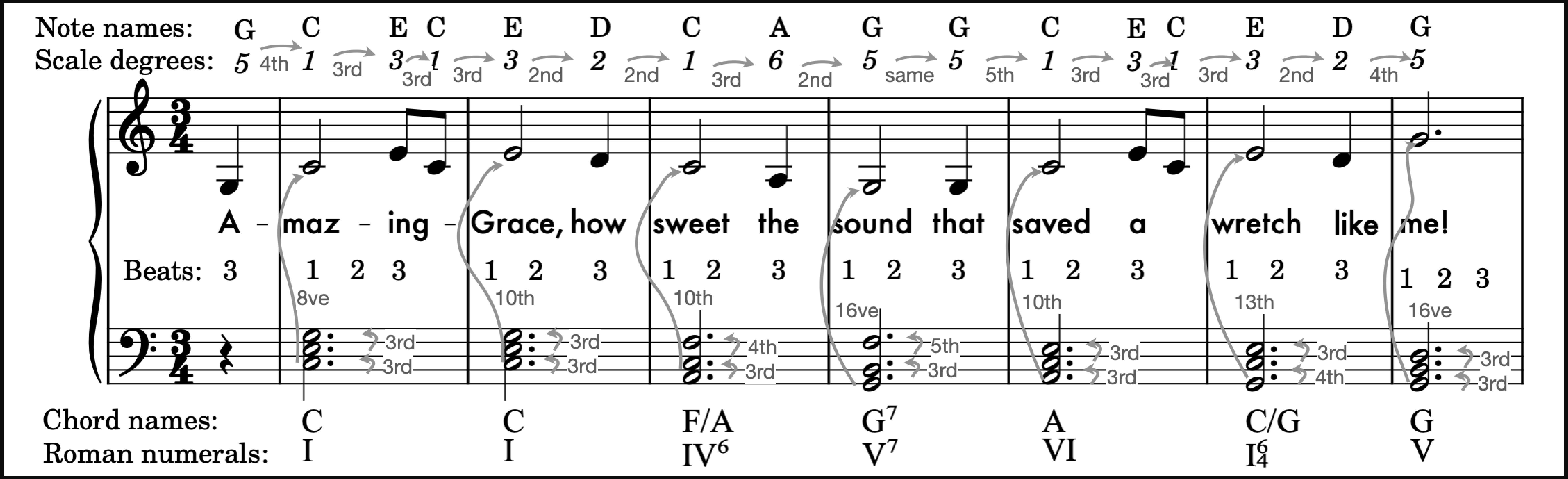
Just think of all the ways you can talk about what's going on here



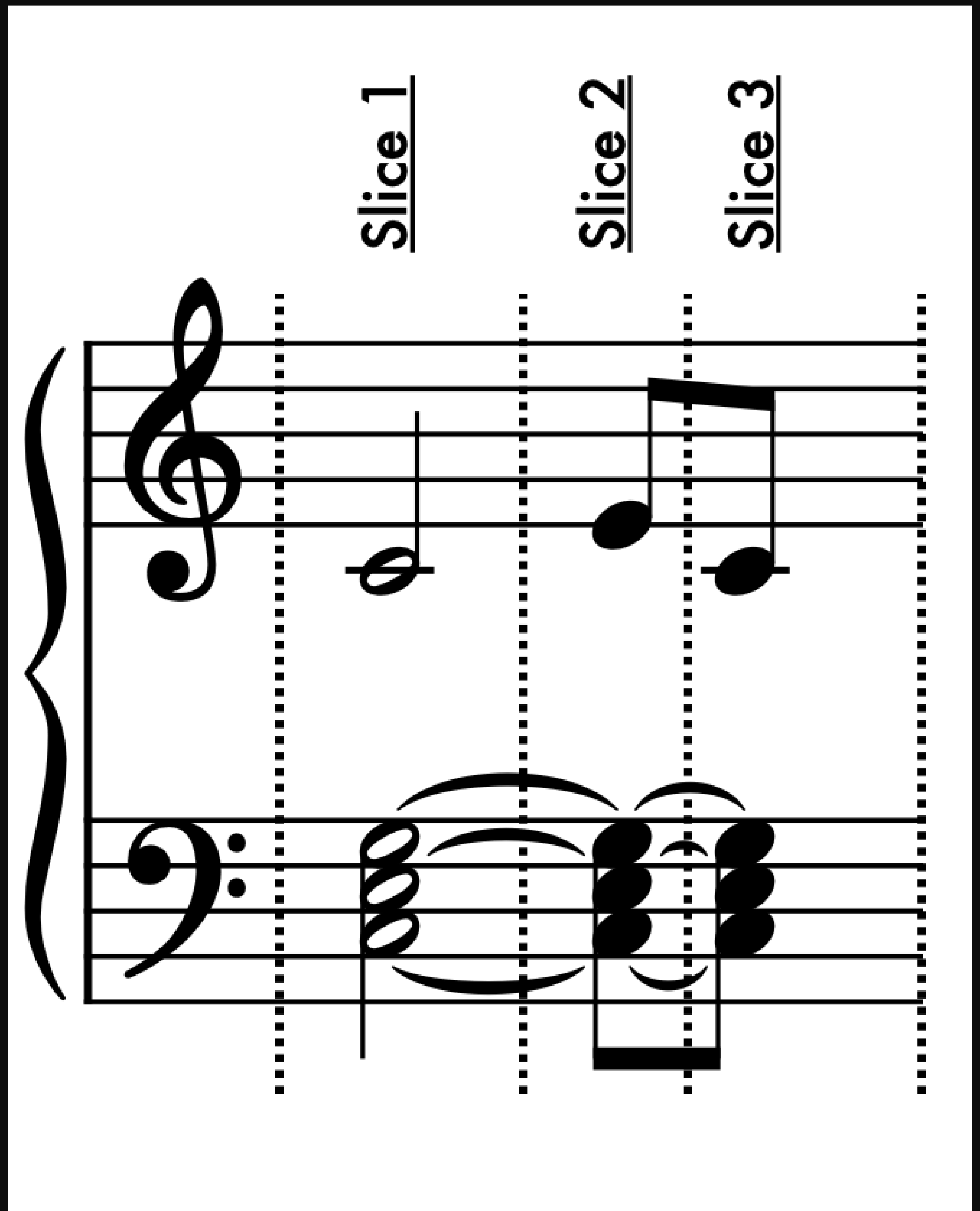
Just think of all the ways you can talk about what's going on here



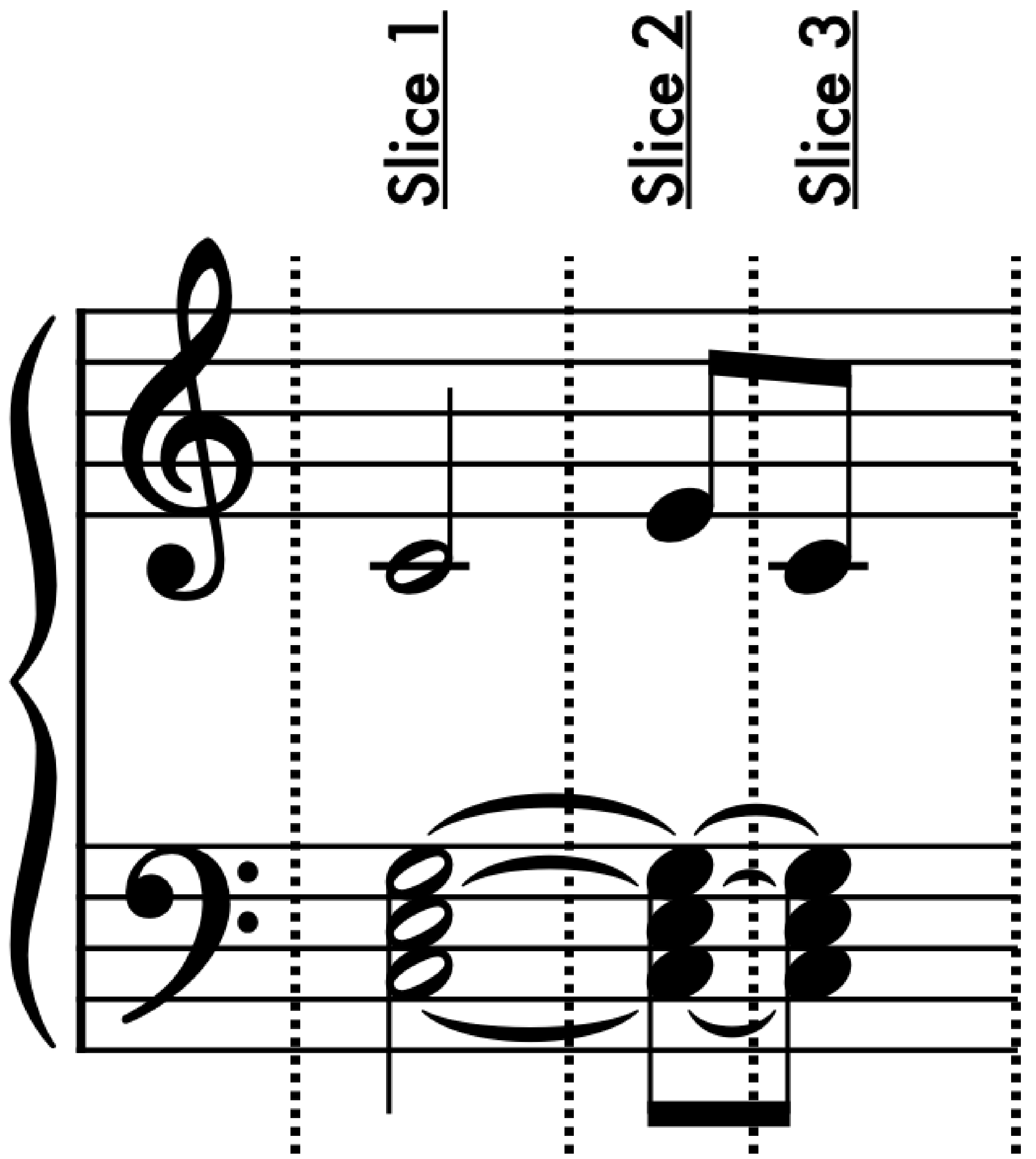
Just think of all the ways you can talk about what's going on here



So, how specifically do we represent these three moments?



So, how specifically do we represent these three moments?

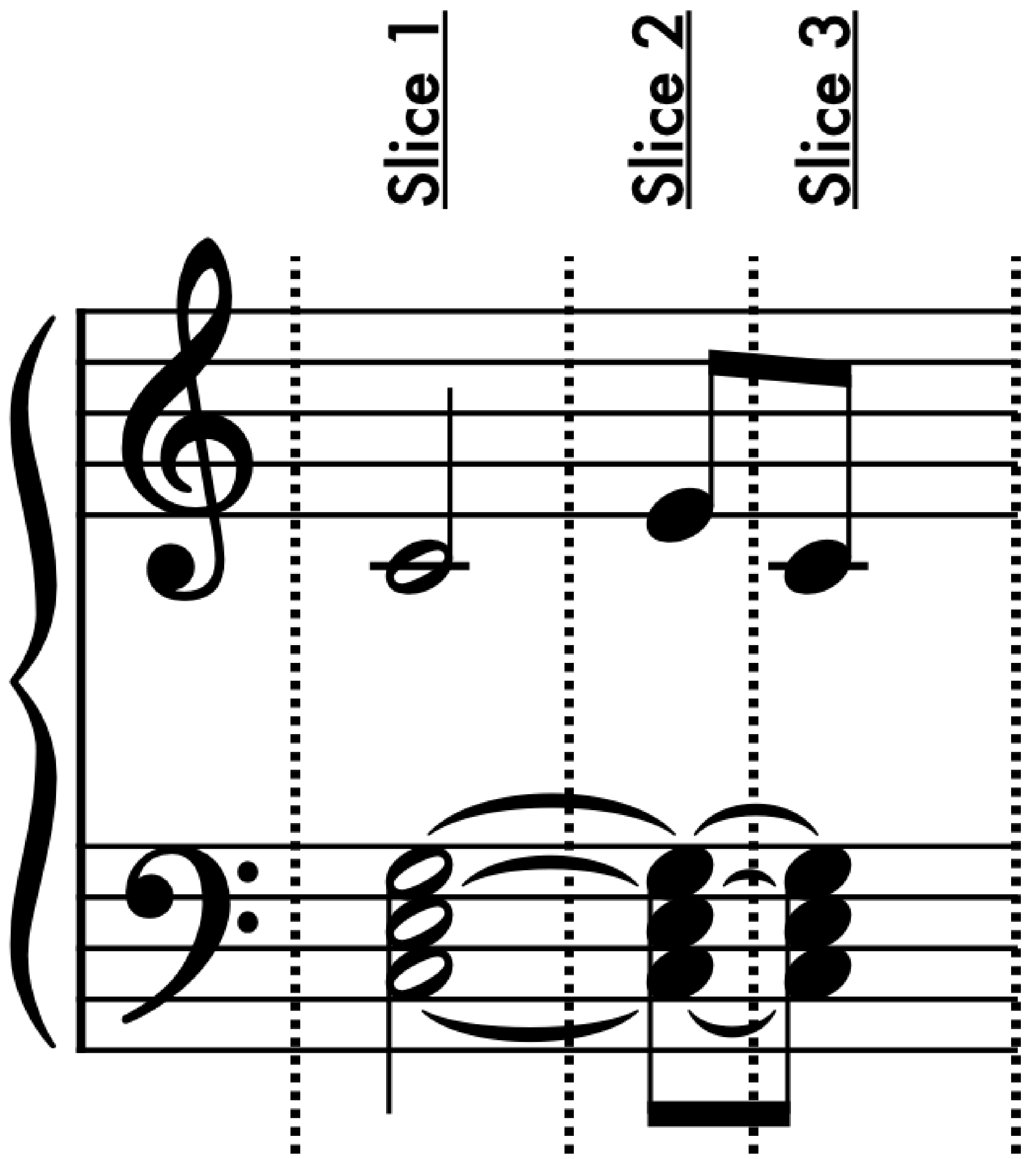


Harmony/melody

Pitches:

{Low C, Low E, Low G, Mid C}
{Low C, Low E, Low G, Mid E}
{Low C, Low E, Low G, Mid C}

So, how specifically do we represent these three moments?



Harmony/melody

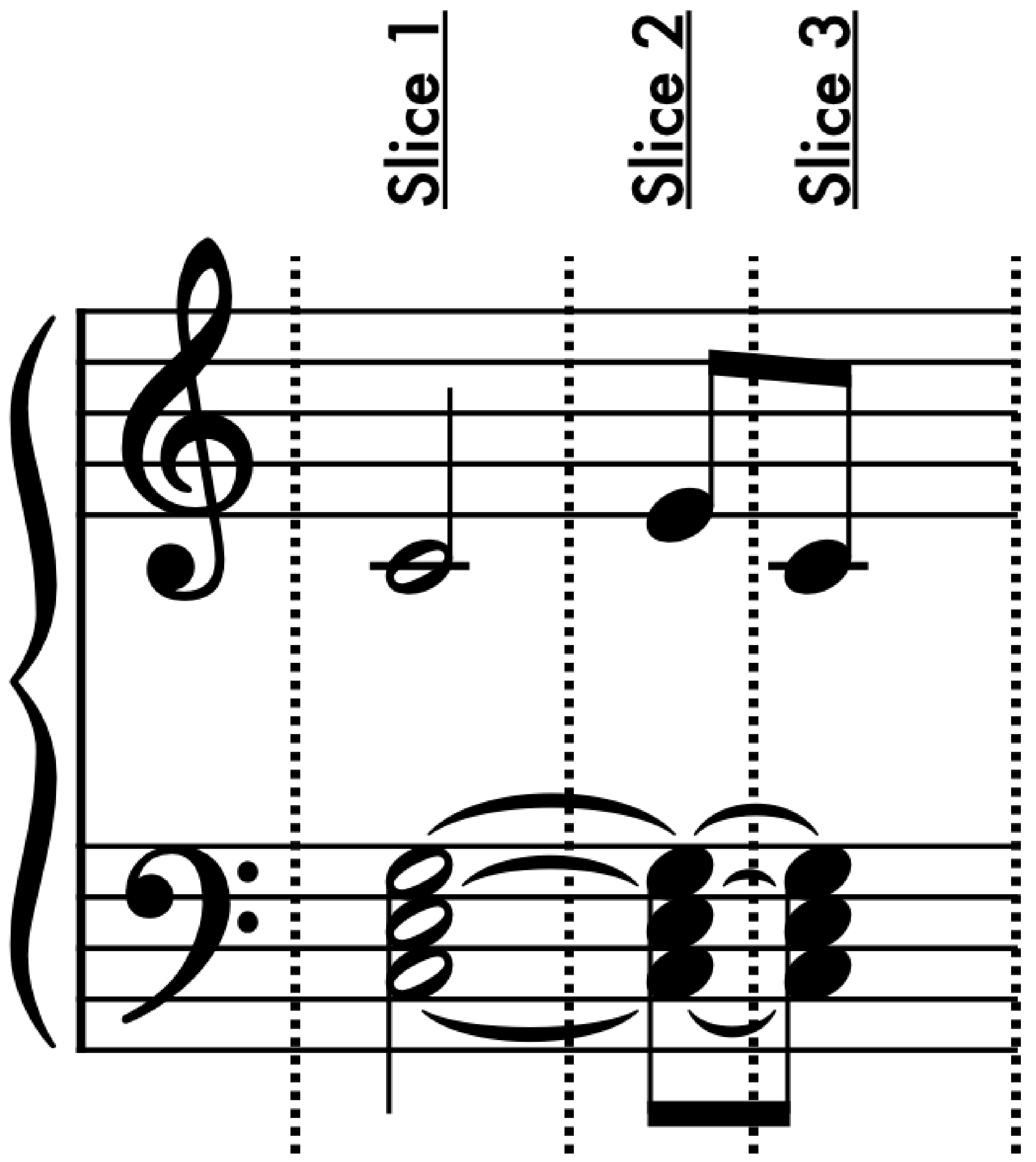
Pitches:

{Low C, Low E, Low G, Mid C}
{Low C, Low E, Low G, Mid E}
{Low C, Low E, Low G, Mid C}

Scale degree & intervals:

{SD 1, & 3rd, 5th, 8th}
{SD 1, & 3rd, 5th, 10th}
{SD 1, & 3rd, 5th, 8th}

So, how specifically do we represent these three moments?



Harmony/melody

<u>Pitches:</u>

{Low C, Low E, Low G, Mid C}
{Low C, Low E, Low G, Mid E}
{Low C, Low E, Low G, Mid C}

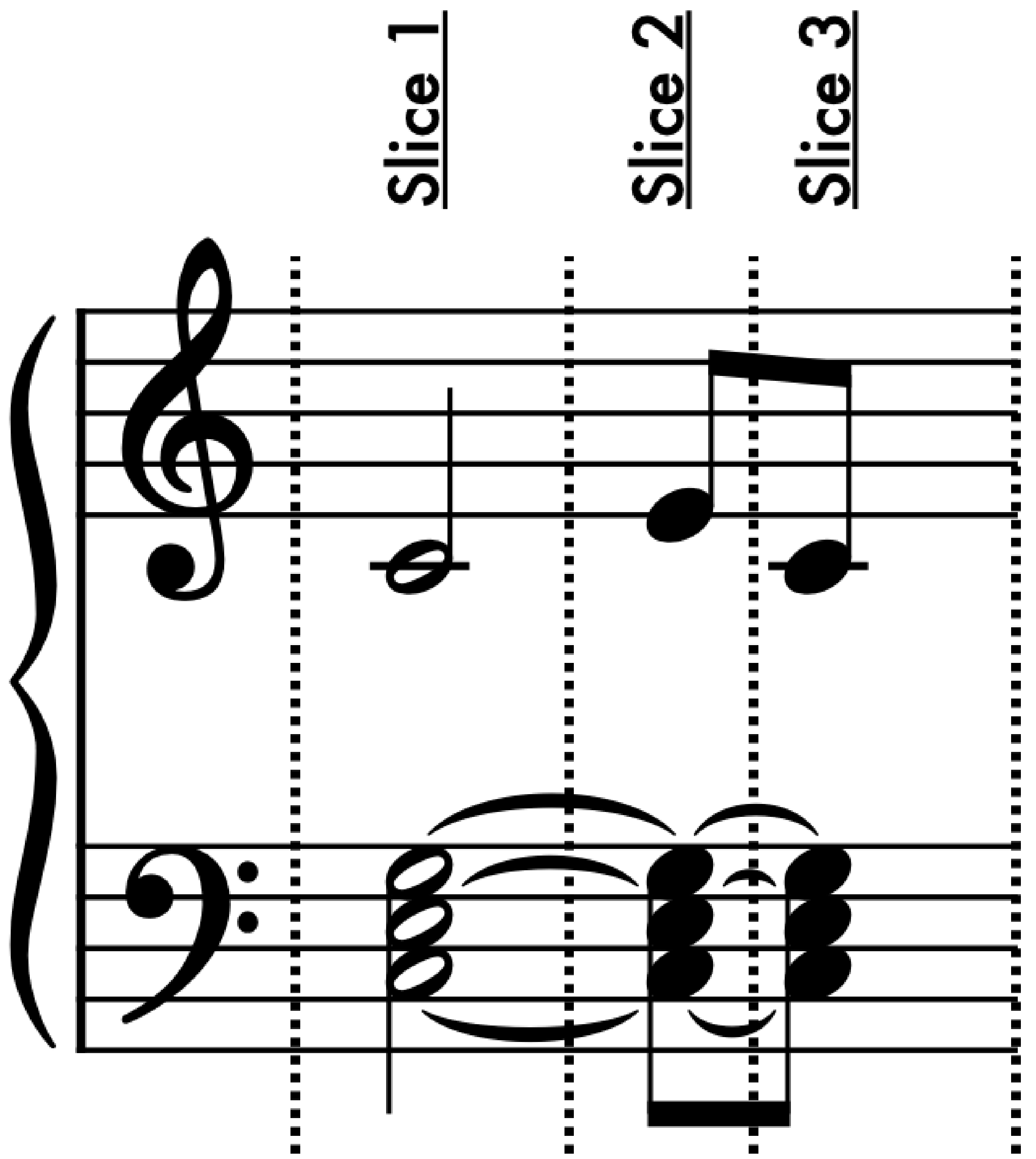
<u>Scale degree & intervals:</u>

{SD 1, & 3rd, 5th, 8th}
{SD 1, & 3rd, 5th, 10th}
{SD 1, & 3rd, 5th, 8th}

<u>Chord types:</u>

{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}

So, how specifically do we represent these three moments?



Harmony/melody

Pitches:

{Low C, Low E, Low G, Mid C}
{Low C, Low E, Low G, Mid E}
{Low C, Low E, Low G, Mid C}

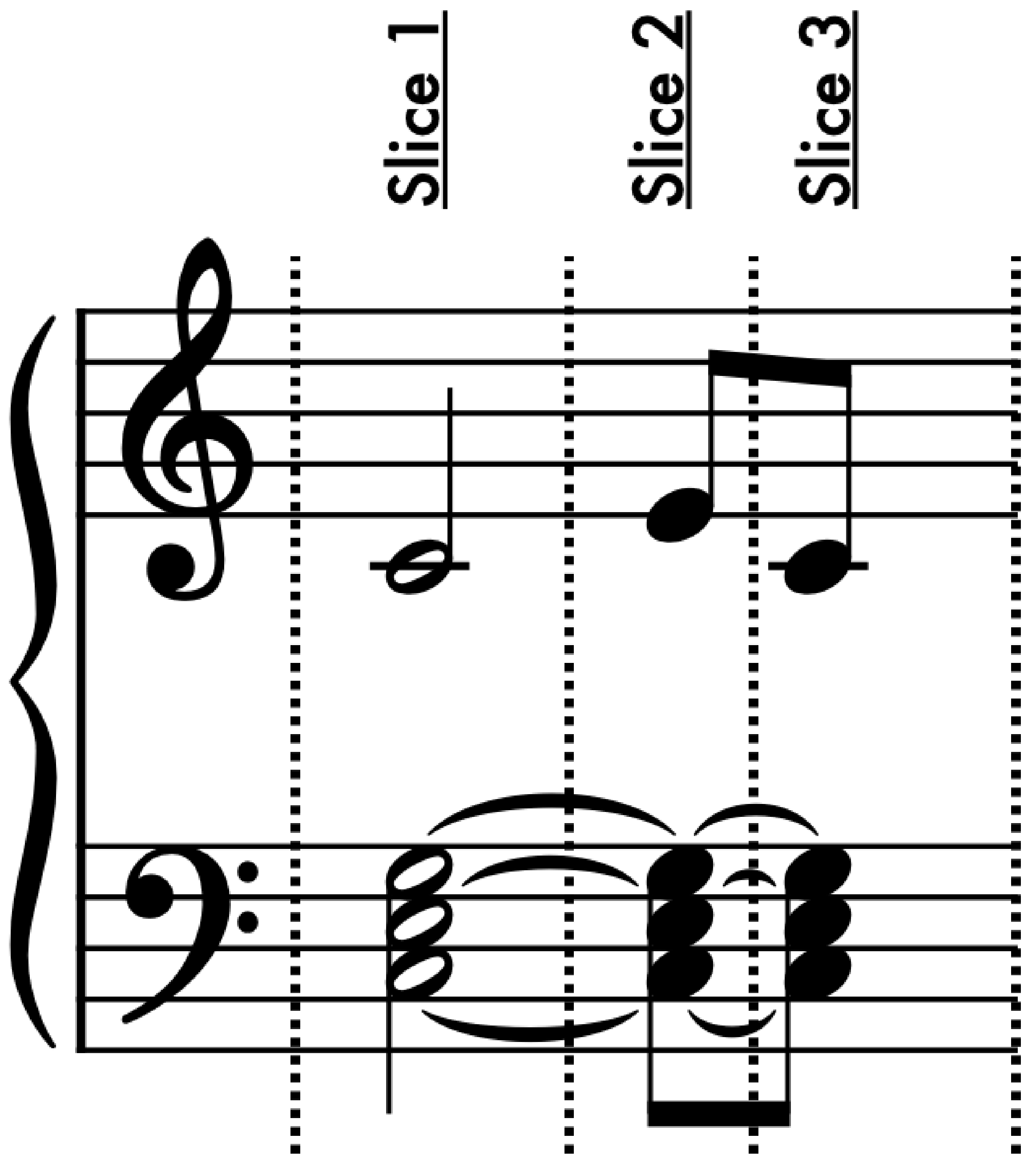
Scale degree & intervals:

{SD 1, & 3rd, 5th, 8th}
{SD 1, & 3rd, 5th, 10th}
{SD 1, & 3rd, 5th, 8th}

Chord types:

{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}

So, how specifically do we represent these three moments?



Harmony/melody

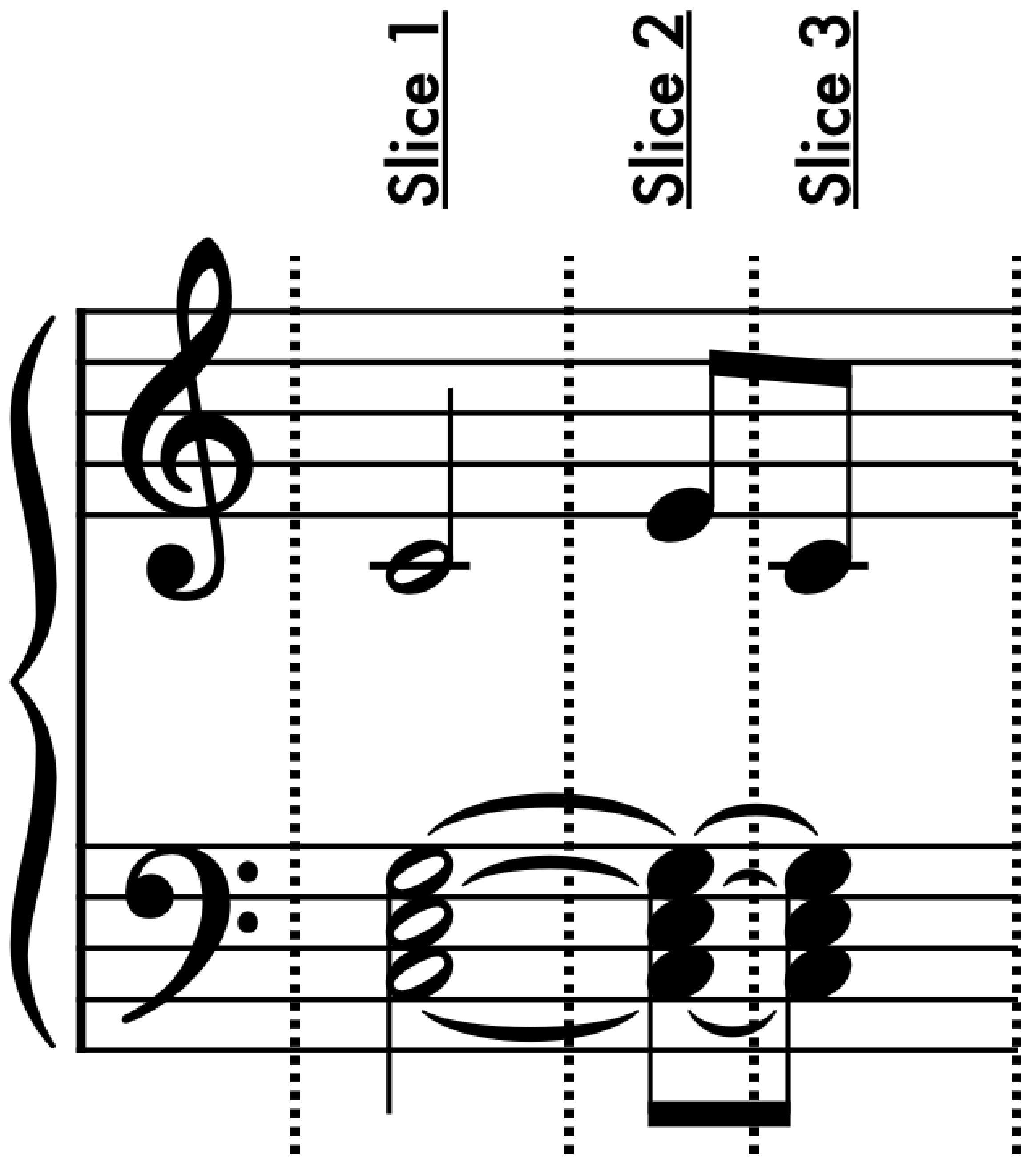
Pitches:
{Low C, Low E, Low G, Mid C}
{Low C, Low E, Low G, Mid E}
{Low C, Low E, Low G, Mid C}

Scale degree & intervals:
{SD 1, & 3rd, 5th, 8th}
{SD 1, & 3rd, 5th, 10th}
{SD 1, & 3rd, 5th, 8th}

Chord types:
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}

Outer voices:
{SD 1 octave above last note, & octave above}
{SD 1 octave above last note, & tenth above}
{SD 1 octave above last note, & octave above}

So, how specifically do we represent these three moments?



Harmony/melody

<u>Pitches:</u>

{Low C, Low E, Low G, Mid C}
{Low C, Low E, Low G, Mid E}
{Low C, Low E, Low G, Mid C}

<u>Scale degree & intervals:</u>

{SD 1, & 3rd, 5th, 8th}
{SD 1, & 3rd, 5th, 10th}
{SD 1, & 3rd, 5th, 8th}

<u>Chord types:</u>

{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}

<u>Outer voices:</u>

{SD 1 octave above last note, & octave above}
{SD 1 octave above last note, & tenth above}
{SD 1 octave above last note, & octave above}

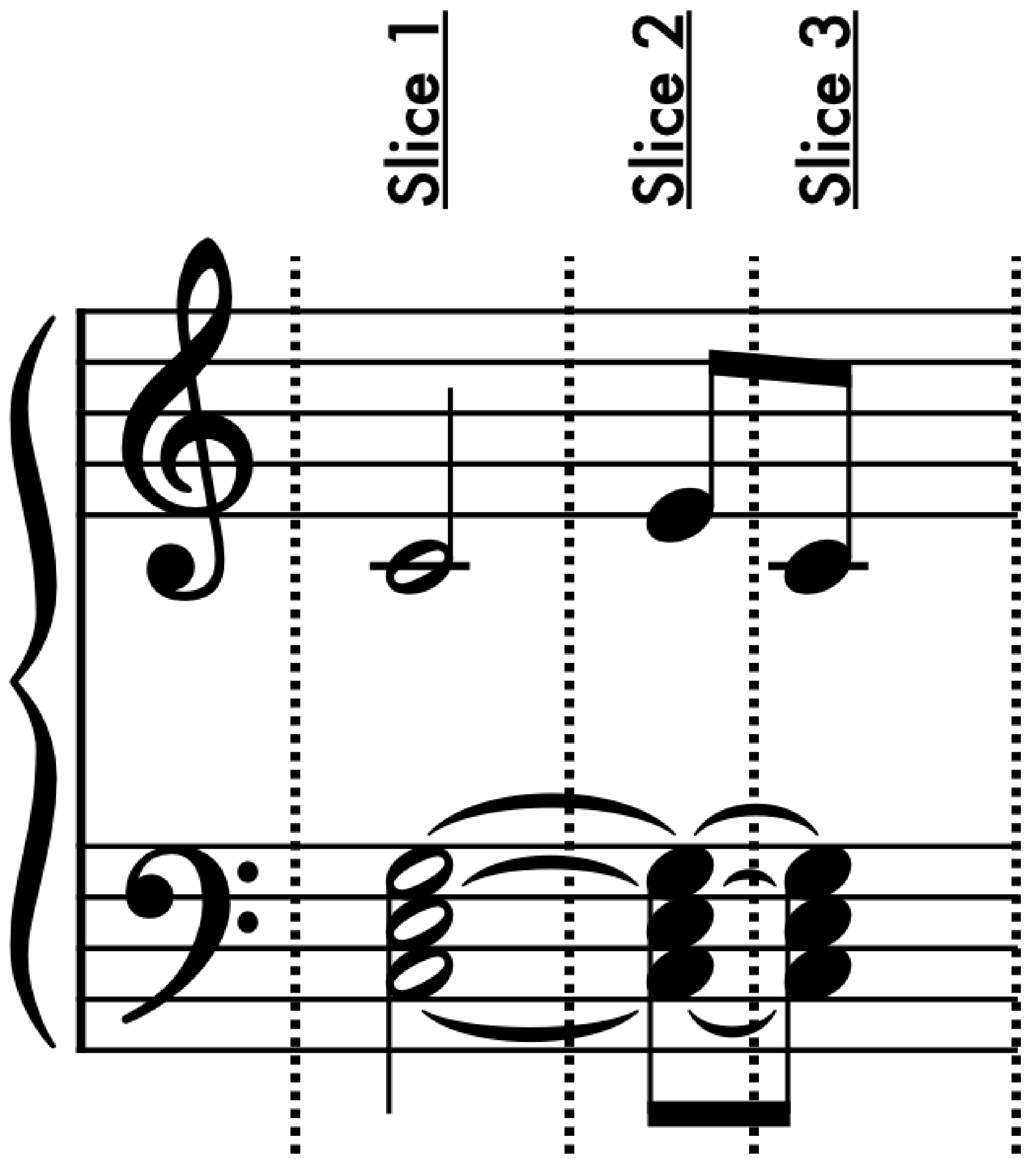
Intervals

<u>Linear intervals</u>

(from previous upbeat Low G)

{↓ 5th, ↓ 3rd, Hold, ↑ 3rd}
{Hold, Hold, Hold, ↑ 3rd }
{Hold, Hold, Hold, ↓ 3rd }

So, how specifically do we represent these three moments?



Harmony/melody

Pitches:
{Low C, Low E, Low G, Mid C}
{Low C, Low E, Low G, Mid E}
{Low C, Low E, Low G, Mid C}

Scale degree & intervals:
{SD 1, & 3rd, 5th, 8th}
{SD 1, & 3rd, 5th, 10th}
{SD 1, & 3rd, 5th, 8th}

Chord types:
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}

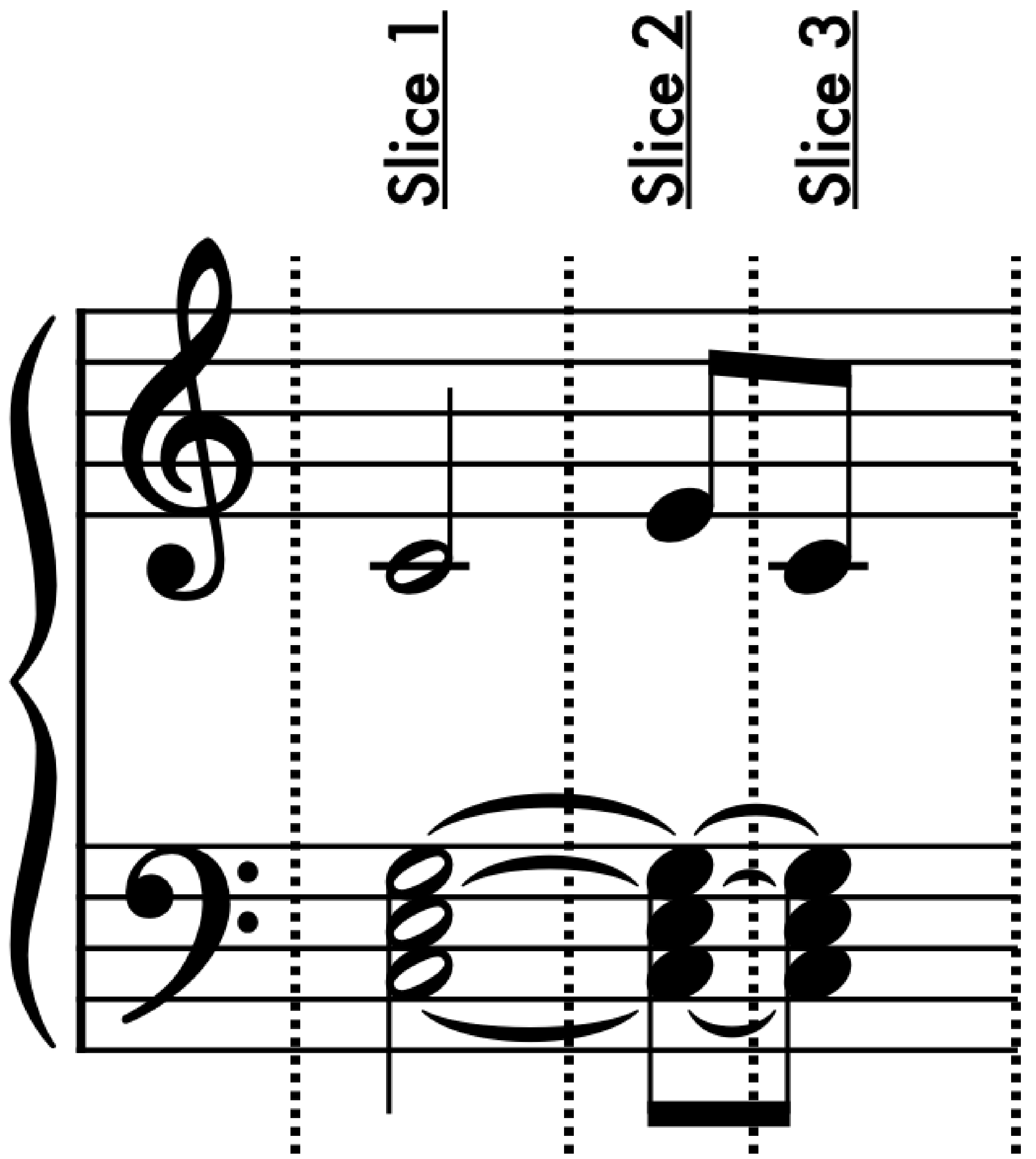
Outer voices:
{SD 1 octave above last note, & octave above}
{SD 1 octave above last note, & tenth above}
{SD 1 octave above last note, & octave above}

Intervals

Linear intervals
(from previous upbeat Low G)
{↓ 5th, ↓ 3rd, Hold, ↑ 3rd}
{Hold, Hold, Hold, ↑ 3rd }
{Hold, Hold, Hold, ↓ 3rd }

Just vertical intervals:
{3rd, 5th, 8th}
{3rd, 5th, 10th}
{3rd, 5th, 8th}

So, how specifically do we represent these three moments?



Harmony/melody

<u>Pitches:</u>
{Low C, Low E, Low G, Mid C}
{Low C, Low E, Low G, Mid E}
{Low C, Low E, Low G, Mid C}

<u>Scale degree & intervals:</u>
{SD 1, & 3rd, 5th, 8th}
{SD 1, & 3rd, 5th, 10th}
{SD 1, & 3rd, 5th, 8th}

<u>Chord types:</u>
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}

<u>Outer voices:</u>
{SD 1 octave above last note, & octave above}
{SD 1 octave above last note, & tenth above}
{SD 1 octave above last note, & octave above}

Intervals

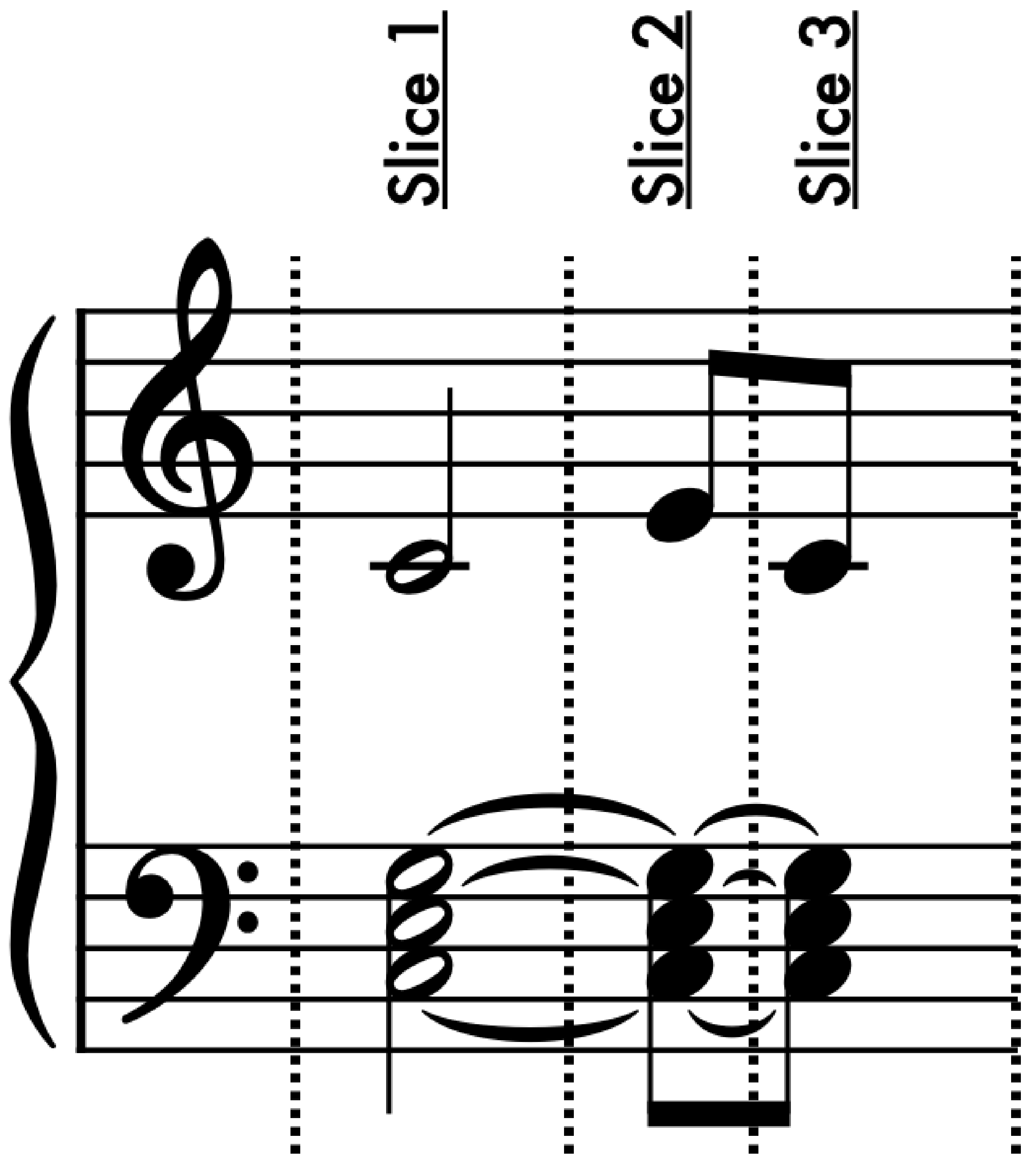
<u>Linear intervals</u>
(from previous upbeat Low G)
{↓ 5th, ↓ 3rd, Hold, ↑ 3rd}
{Hold, Hold, Hold, ↑ 3rd }
{Hold, Hold, Hold, ↓ 3rd }

<u>Just vertical intervals:</u>
{3rd, 5th, 8th}
{3rd, 5th, 10th}
{3rd, 5th, 8th}

Meter/rhythm

<u>Duration & Metric Position:</u>
{2 beats on downbeat}
{.5 beat on beat 3}
{.5 beat on beat 3.5}

So, how specifically do we represent these three moments?



Harmony/melody

Pitches:
{Low C, Low E, Low G, Mid C}
{Low C, Low E, Low G, Mid E}
{Low C, Low E, Low G, Mid C}

Scale degree & intervals:
{SD 1, & 3rd, 5th, 8th}
{SD 1, & 3rd, 5th, 10th}
{SD 1, & 3rd, 5th, 8th}

Chord types:
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}

Outer voices:
{SD 1 octave above last note, & octave above}
{SD 1 octave above last note, & tenth above}
{SD 1 octave above last note, & octave above}

Intervals

Linear intervals
(from previous upbeat Low G)
{↓ 5th, ↓ 3rd, Hold, ↑ 3rd}
{Hold, Hold, Hold, ↑ 3rd }
{Hold, Hold, Hold, ↓ 3rd }

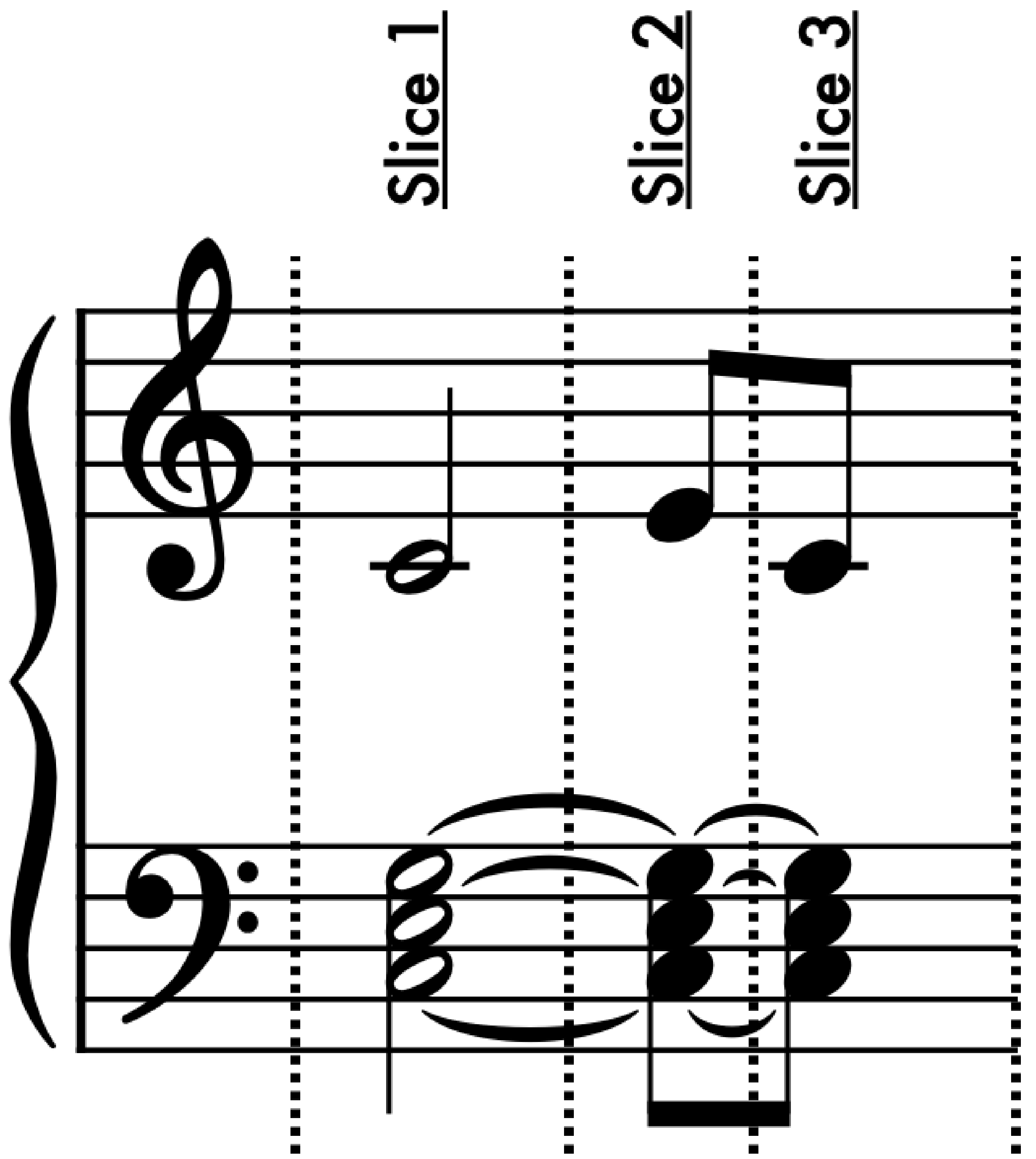
Just vertical intervals:
{3rd, 5th, 8th}
{3rd, 5th, 10th}
{3rd, 5th, 8th}

Meter/rhythm

Duration & Metric Position:
{2 beats on downbeat}
{.5 beat on beat 3}
{.5 beat on beat 3.5}

Duration:
{2 beats}
{.5 beat}
{.5 beat}

So, how specifically do we represent these three moments?



Harmony/melody

<u>Pitches:</u>

{Low C, Low E, Low G, Mid C}
{Low C, Low E, Low G, Mid E}
{Low C, Low E, Low G, Mid C}

<u>Scale degree & intervals:</u>

{SD 1, & 3rd, 5th, 8th}
{SD 1, & 3rd, 5th, 10th}
{SD 1, & 3rd, 5th, 8th}

<u>Chord types:</u>

{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}

<u>Outer voices:</u>

{SD 1 octave above last note, & octave above}
{SD 1 octave above last note, & tenth above}
{SD 1 octave above last note, & octave above}

Intervals

<u>Linear intervals</u>

(from previous upbeat Low G)

{↓ 5th, ↓ 3rd, Hold, ↑ 3rd}
{Hold, Hold, Hold, ↑ 3rd }
{Hold, Hold, Hold, ↓ 3rd }

<u>Just vertical intervals:</u>

{3rd, 5th, 8th}
{3rd, 5th, 10th}
{3rd, 5th, 8th}

Meter/rhythm

<u>Duration & Metric Position:</u>

{2 beats on downbeat}
{.5 beat on beat 3}
{.5 beat on beat 3.5}

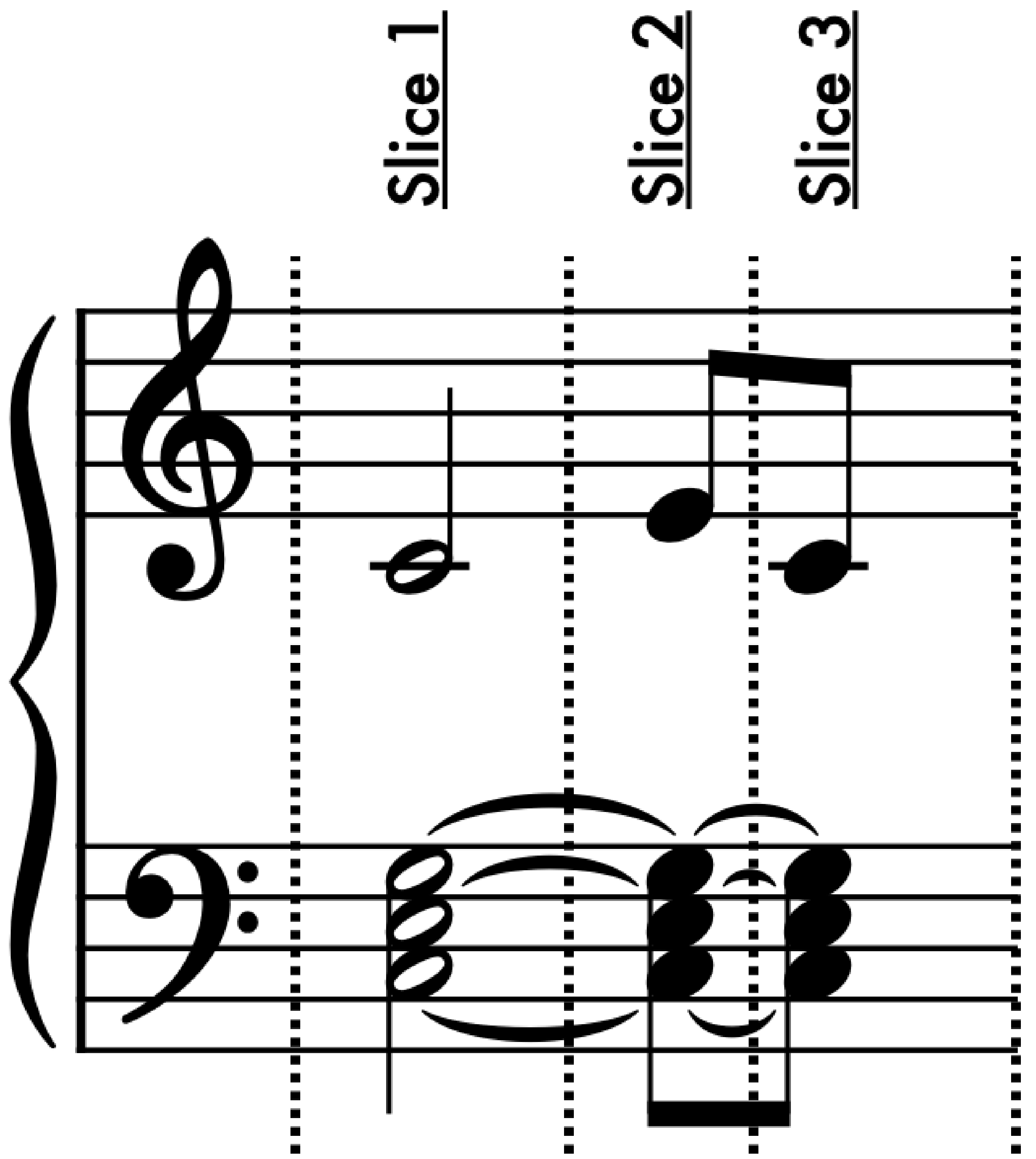
<u>Duration:</u>

{2 beats}
{.5 beat}
{.5 beat}

<u>Position in measure:</u>

{Downbeat}
{Beat 3}
{Beat 3.5}

So, how specifically do we represent these three moments?



Harmony/melody

Pitches:
{Low C, Low E, Low G, Mid C}
{Low C, Low E, Low G, Mid E}
{Low C, Low E, Low G, Mid C}

Scale degree & intervals:
{SD 1, & 3rd, 5th, 8th}
{SD 1, & 3rd, 5th, 10th}
{SD 1, & 3rd, 5th, 8th}

Chord types:
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}
{SD 1, 3, 5 w- SD 1 lowest}

Outer voices:
{SD 1 octave above last note, & octave above}
{SD 1 octave above last note, & tenth above}
{SD 1 octave above last note, & octave above}

Intervals

Linear intervals
(from previous upbeat Low G)
{↓ 5th, ↓ 3rd, Hold, ↑ 3rd}
{Hold, Hold, Hold, ↑ 3rd }
{Hold, Hold, Hold, ↓ 3rd }

Just vertical intervals:
{3rd, 5th, 8th}
{3rd, 5th, 10th}
{3rd, 5th, 8th}

Meter/rhythm

Duration & Metric Position:
{2 beats on downbeat}
{.5 beat on beat 3}
{.5 beat on beat 3.5}

Duration:
{2 beats}
{.5 beat}
{.5 beat}

Position in measure:
{Downbeat}
{Beat 3}
{Beat 3.5}

Beat strength:
{strong}
{weak}
{weaker}

Representation

- It's just not obvious what data engineers should be feeding to a deep learning model
- This means that many people are trying many different approaches
- And this spreads musical AI's already-limited resources thin, as different teams try out many different solutions

Force #4: Structure

Motivation	Examples	Representation	Structure
Why are people making models of musical AI?	What datasets are people using to train and test their AI models?	How are programmers representing musical events in their AI?	What aspects of musical organization are being learned by the AI?

Force #4: Structure

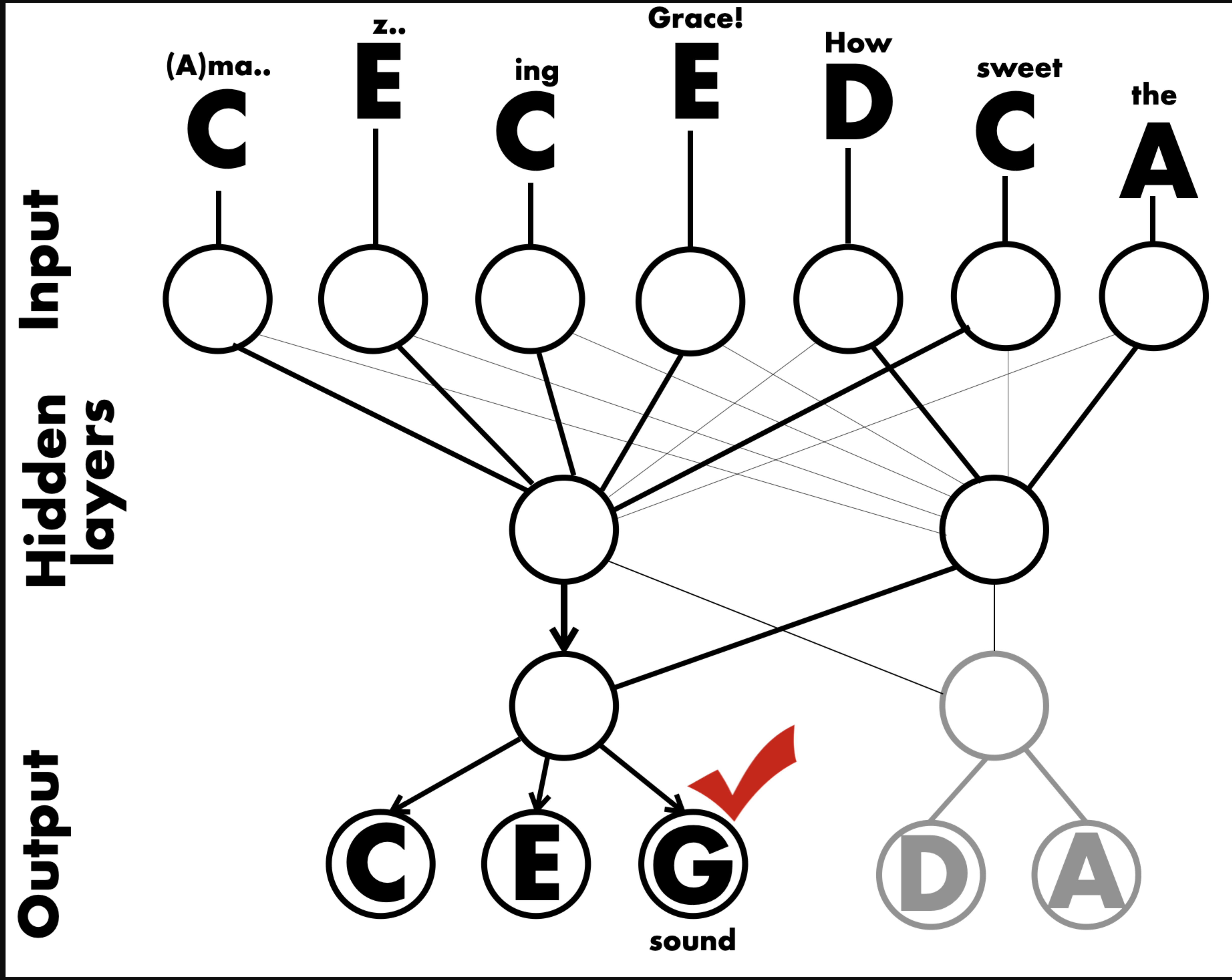
Motivation	Examples	Representation	Structure
Why are people making models of musical AI?	What datasets are people using to train and test their AI models?	How are programmers representing musical events in their AI?	What aspects of musical organization are being learned by the AI?

Force #4: Structure

Music is constructed in complex ways that are hard for an AI to learn

Structure

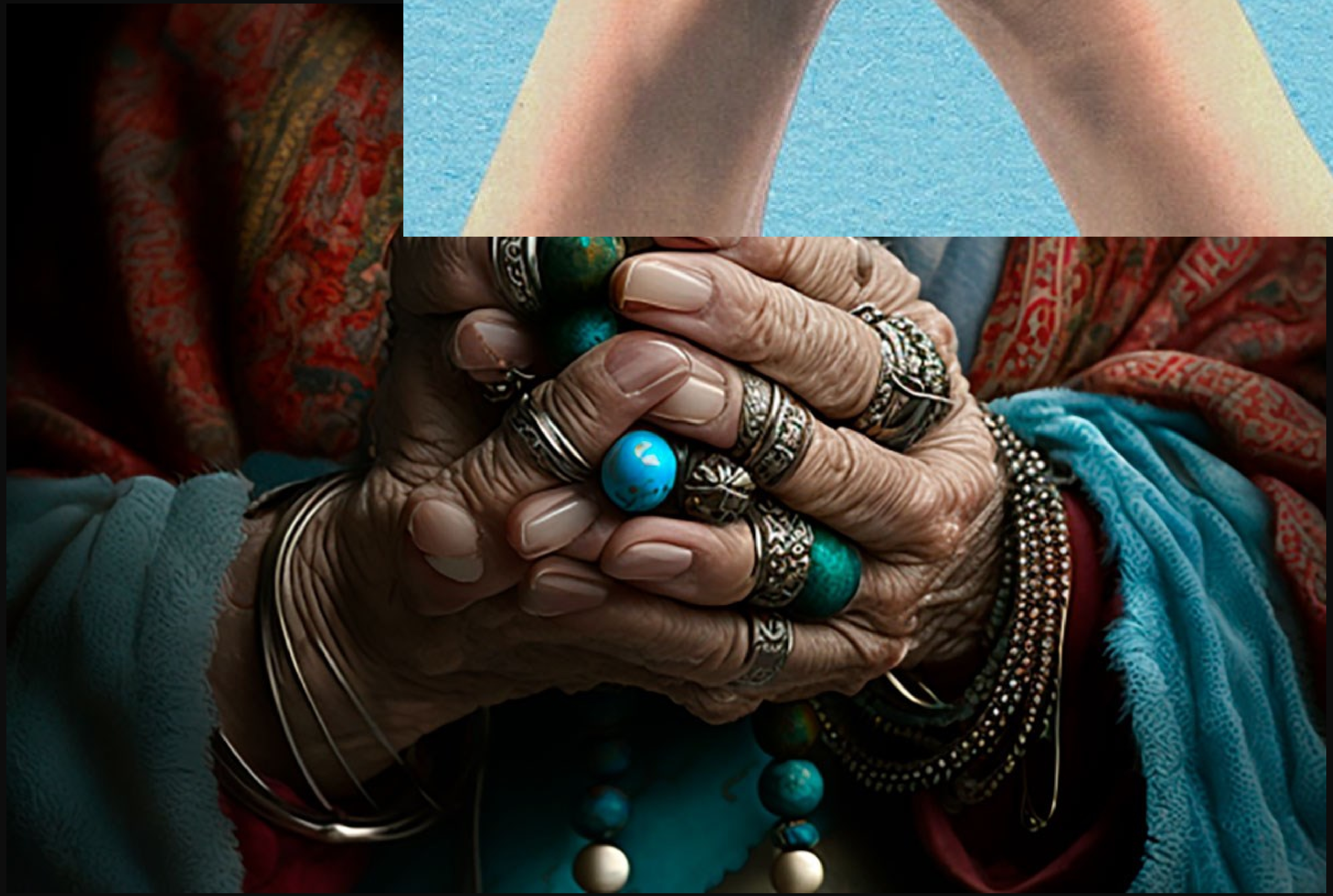
- Deep Learning groups its data based on patterns and regularities
- This means it's very good at learning nested determined proximities:
  - Things which are near to each other (proximities)
  - Things that occur a lot (they're determined)
  - Things that it can chunk (they're nested)



Structure

- Neural networks have a hard time with things that:
- can take many forms (aren't determined)
- are spread out (aren't proximate), and
- Have independent components (aren't nested)
- Like hands…

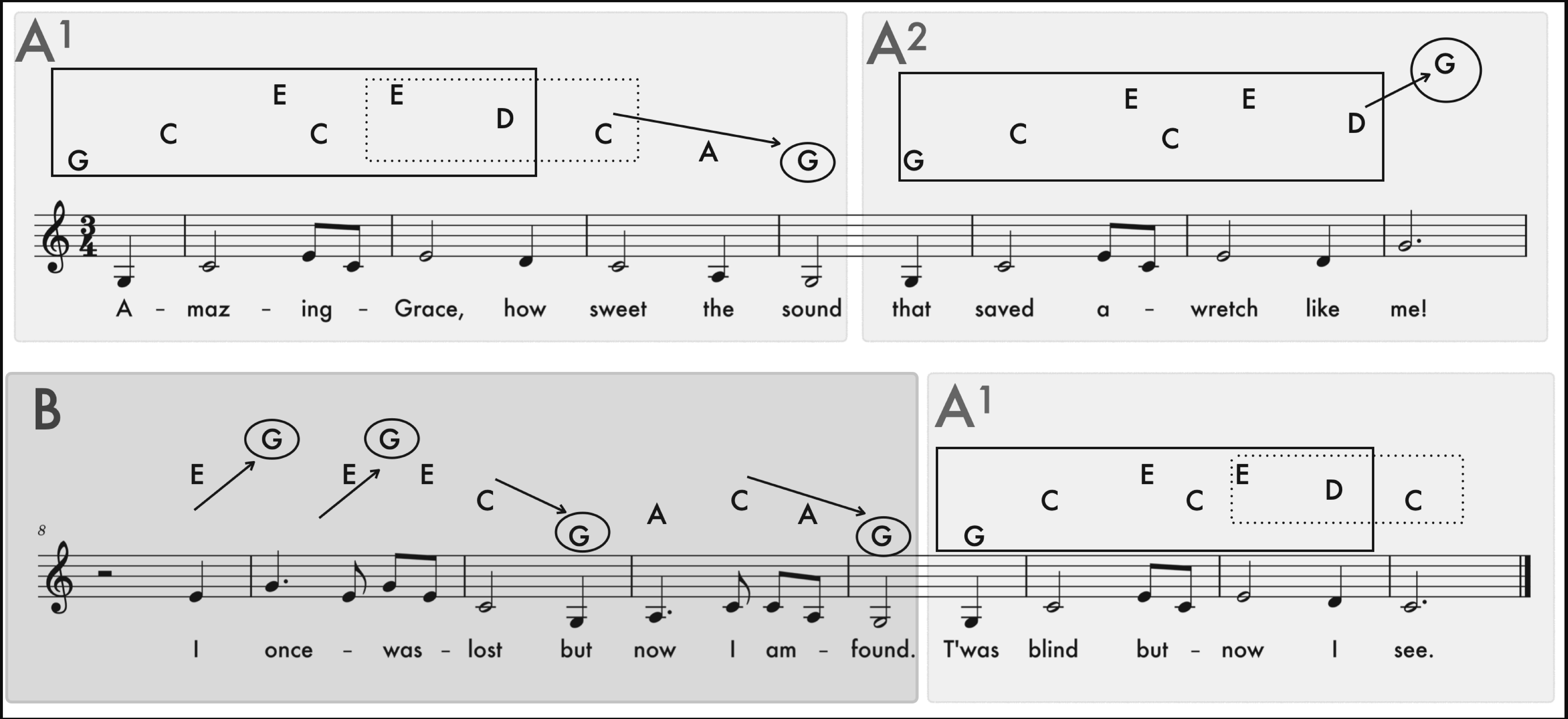


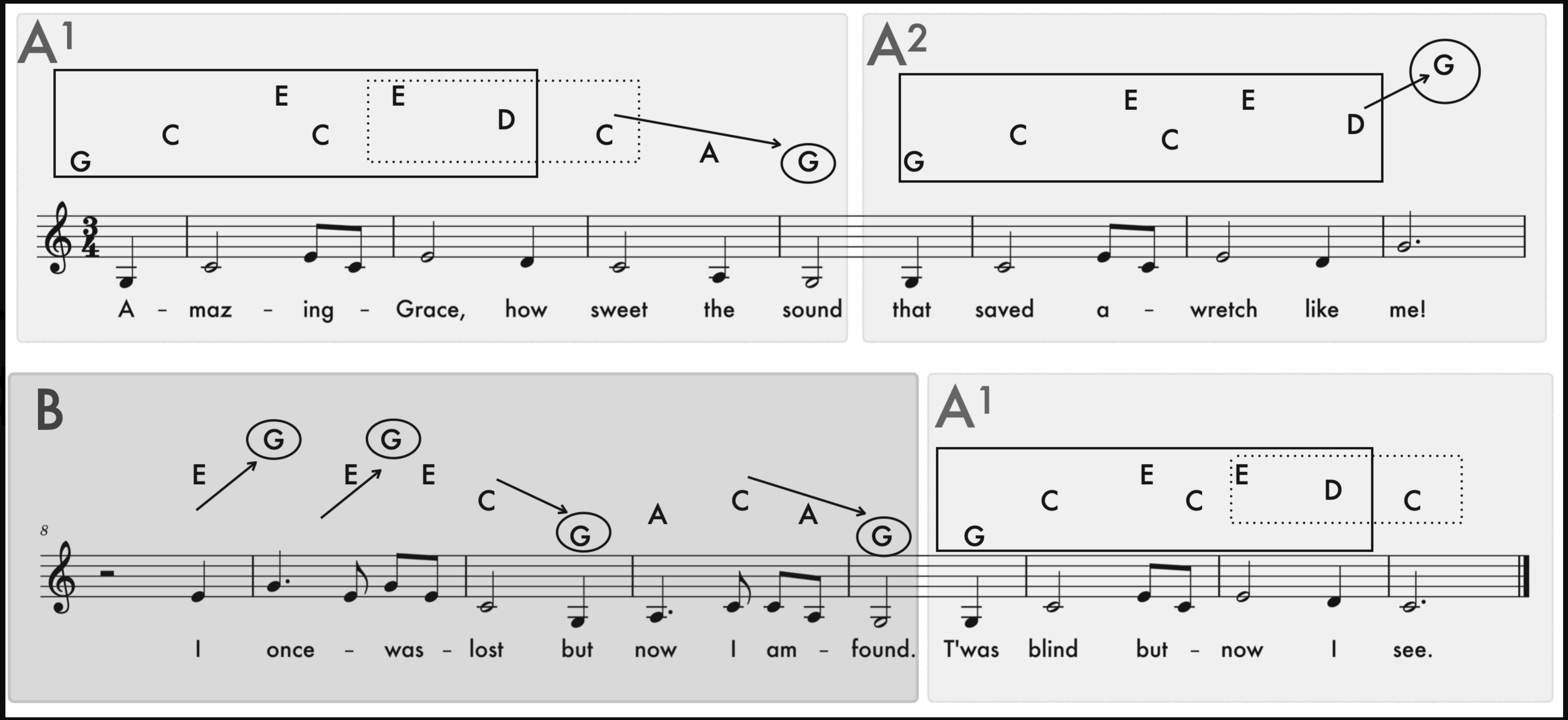
Musical structure is like hands

Musical connections are spread apart, occur in many different ways, and don’t always combine into some predictable whole


A1
G C E C E D C A G
A - maz - ing - Grace, how sweet the sound
A2
G C E C E D G
that saved a - wretch like me!
B
8
E G E G E C G A C A G
I once - was - lost but now I am - found.
A1
G C E C E D C
T'was blind but - now I see.



A1
G C E C E D C A G
A - maz - ing - Grace, how sweet the sound
A2
G C E C E D G
that saved a - wretch like me!
B
8
E G E G E C G A C A G
I once - was - lost but now I am - found.
A1
G C E C E D C
T'was blind but - now I see.



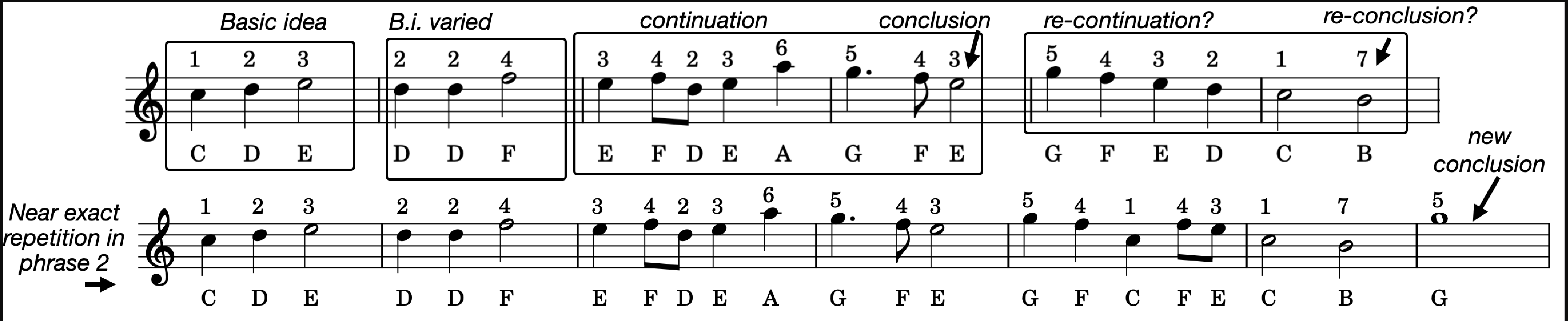
- All of this makes musical sense and contribute to a flowing and coherent tune
- But none of it was absolutely determined to occur, much of it is spread over larger non-proximate swaths of time, and while some of the music can chunk together, not all of its components nest neatly.

Structure: State of the Art AI

- This example is a “Protestant Christian Hymn Tune” from Udio.

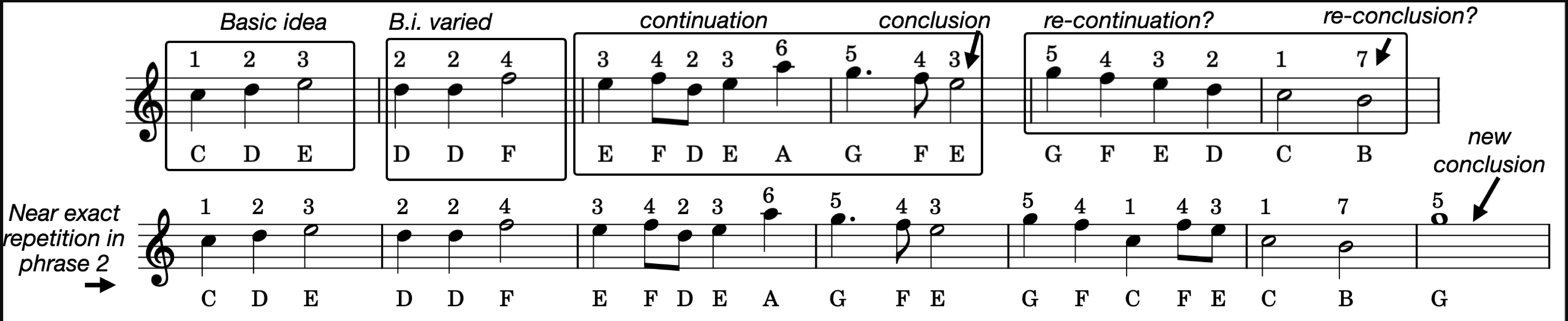
Structure: State of the Art AI

- This example is a "Protestant Christian Hymn Tune" from Udio.



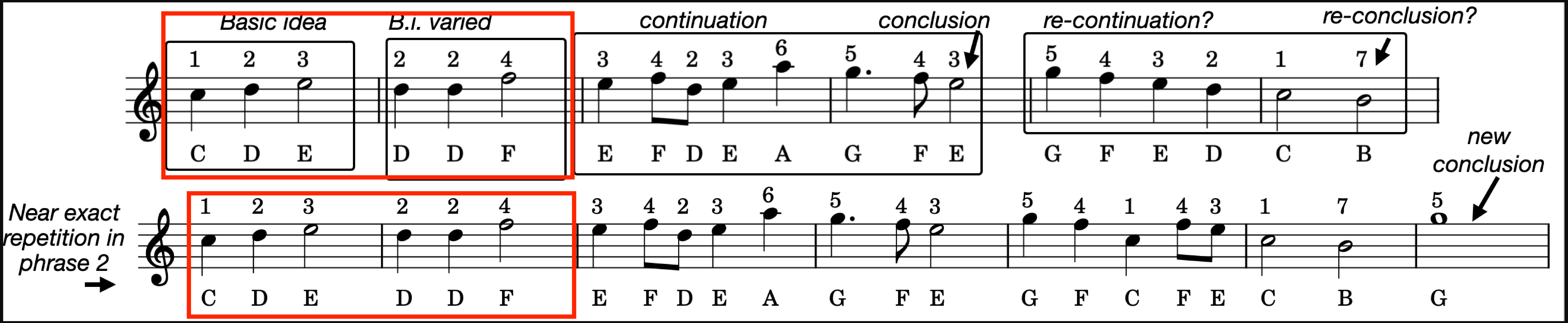
Structure: State of the Art AI

- It's got some nice elements…



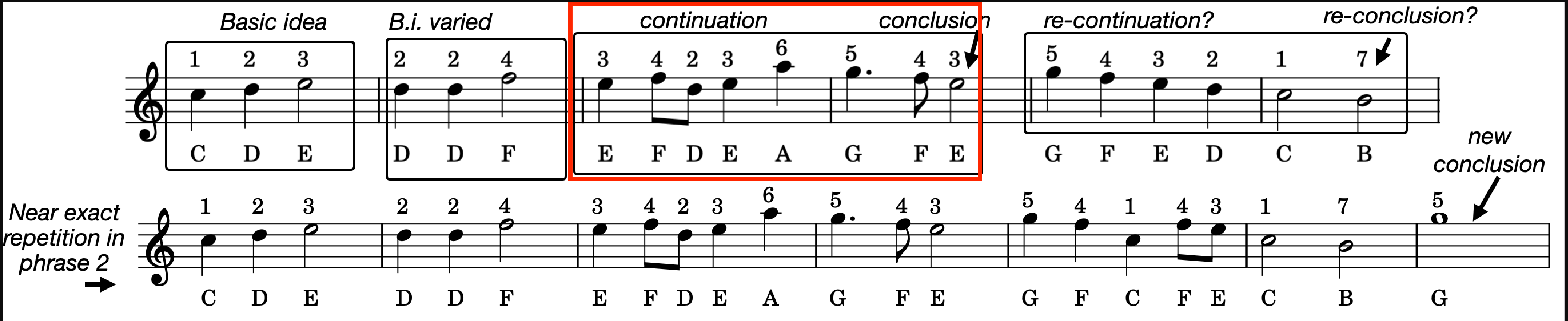
Structure: State of the Art AI

- It’s got some nice elements…



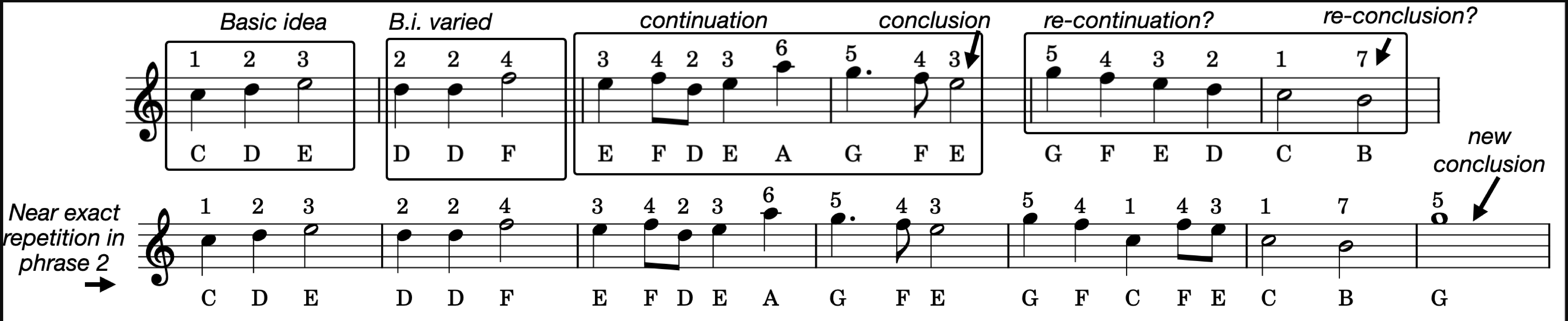
Structure: State of the Art AI

- It's got some nice elements…



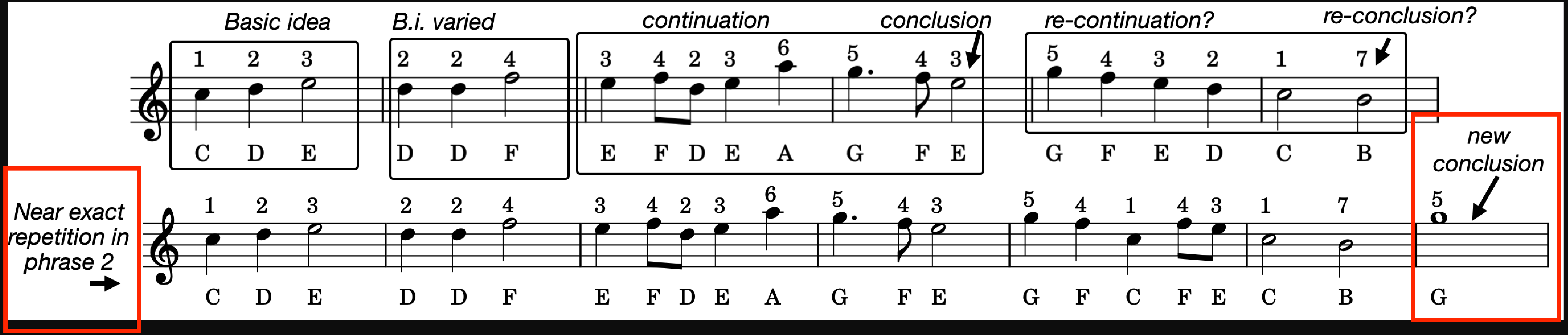
Structure: State of the Art AI

- Its structure is very stilted to try to cram in nested determined proximities



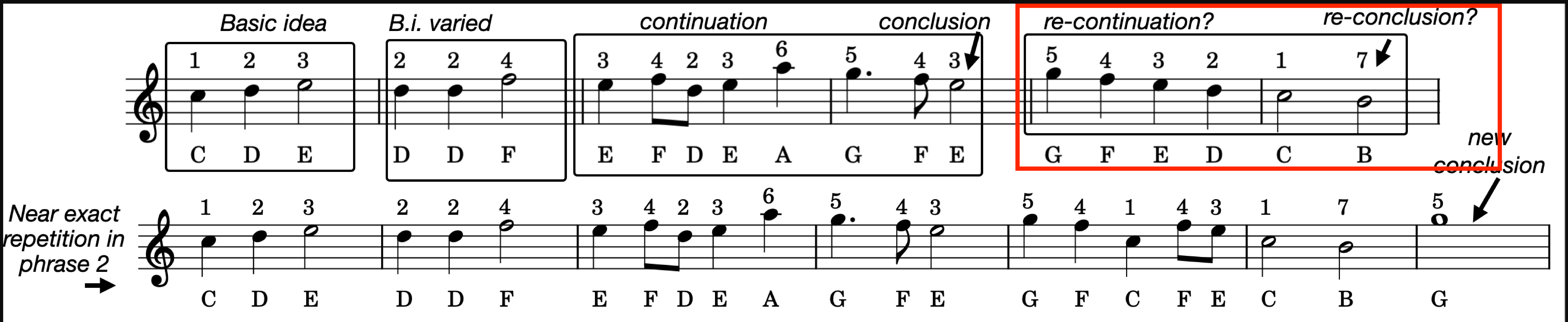
Structure: State of the Art AI

- Its structure is very stilted to try to cram in nested determined proximities



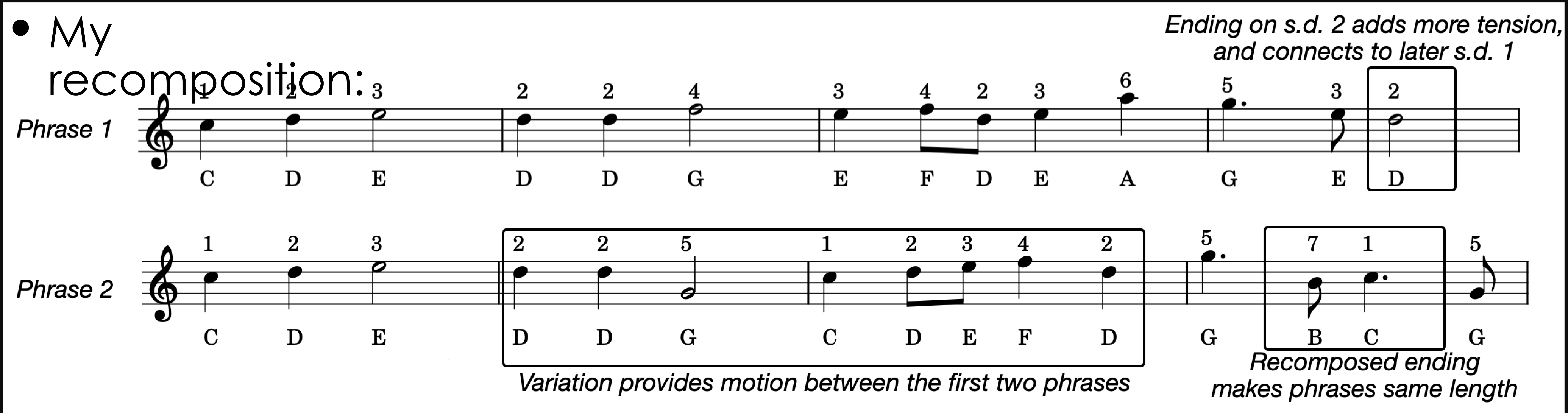
Structure: State of the Art AI

- Its structure is very stilted to try to cram in nested determined proximities



Structure: State of the Art AI

- My recomposition:



Force #5: Interpretation

Motivation	Examples	Representation	Structure	Interpretation
Why are people making models of musical AI?	What datasets are people using to train and test their AI models?	How are programmers representing musical events in their AI?	What aspects of musical organization are being learned by the AI?	What value are listeners drawing from music generated by AI?

Force #5: Interpretation

Motivation	Examples	Representation	Structure	Interpretation
Why are people making models of musical AI?	What datasets are people using to train and test their AI models?	How are programmers representing musical events in their AI?	What aspects of musical organization are being learned by the AI?	What value are listeners drawing from music generated by AI?

Force #5: Interpretation

We might really only like music we know was created by other humans

Interpretation

When describing the designs for the very first computer in 1842,

Ada King, Countess of Lovelace,

wrote that she believed that Artificial Intelligence would never be able to create original content because it's always simply recombining content that's fed into it.

Interpretation

In 1949, when describing the new-fangled calculating machines that helped win the war,

the neurosurgeon and professor George Jefferson wrote in a medical journal that the “mind of the mechanical man” would never…

“write a sonnet or compose a concerto because of thoughts and emotions felt… no mechanism could feel (and not merely artificially signal…) grief… be warmed by flattery, be made miserable… be angry or depressed when it cannot get what it wants.”

Interpretation

- All of this is arguable, but it comes down to two big issues, the symbol grounding problem and the embodiment problem
    - And both are particularly pronounced in music

Symbol grounding

- Meaning arises from a combination of:
  - experiencing something and connecting an image, word, or gesture to that experience (Experiential Knowledge)
  - and connecting those symbols to one another, like reading about some event, attributing some description to music, or interpreting a poem (Associational Knowledge)

Symbol grounding

- Because Artificial Intelligence learns from identifying patterns in a dataset,
  - It becomes an expert in all the contexts in which the words “happy” or “cold” are associated — it knows where these words are appropriate
  - But it will never experience what is means to be “cold” or “happy”
  - It’s reasonable to be very skeptical about whether the computer understands temperature or happiness because it learns only through identifying patterns, connections, and orderings—association

Embodiment

- And in order to experience the world, you need a body
  - Human emotions, perceptions, understanding, cognition, are all mediated through the human body
  - In order to experience the world, that experience needs to be fleshy
  - It's reasonable to be very skeptical about whether a silicon-based intelligence can ever actually experience that world

Music

- We particularly value embodied experiential content in music
- Music doesn't communicate information. Instead, music is a way for humans to socially connect, and to embed and share experiences
  - I'll spare you the philosophy and cognition behind this claim, because this fact is so visceral and easily demonstrated…

Music

- Imagine a bullied, closeted teenager running to their room, and throwing on their headphones.
- Would they ever listen to an AI generated track? No!
    - To them, music is a way to process their own lived experience
    - And it's a way to connect that experience to a shared human experience

Music

- Imagine if these tracks created in the Kendrick/Drake conflict were random tunes generated by AI, not derived from actual human drama
  - We wouldn’t take a second listen!
  - We care about the human stories behind music, and music’s ability to embed lived experiences that we can empathize with

Music

- Imagine you’re a jazz musician of the future, and you’re improvising with an AI on stage
  - What a weird and unfulfilling experience!
  - Part of performing is the thrill of engaging with another musician, looking at them, and conversing and collaborating with them in the moment
  - Removing the human removes the fun!

Music

- Imagine using an AI-generated hymn for your parent's funeral.
- This would feel very inappropriate because…
    - This music encapsulates the lived experience of grief
    - This music connects us as a community to that shared experience

Interpretation

- AI music will simply be interpreted differently than human music
  - As AI grows in sophistication, I believe we're going to increasingly realize how important the fact that a human made it is important to our enjoyment of music
  - We're going to want to know that a human (with the same experiential and embodied life as ours) is on the other end of our music, and we'll want to search for the particular assurance

The river streams into itself

Motivation	Examples	Representation	Structure	Interpretation
Why are people making models of musical AI?	What datasets are people using to train and test their AI models?	How are programmers representing musical events in their AI?	What aspects of musical organization are being learned by the AI?	What value are listeners drawing from music generated by AI?

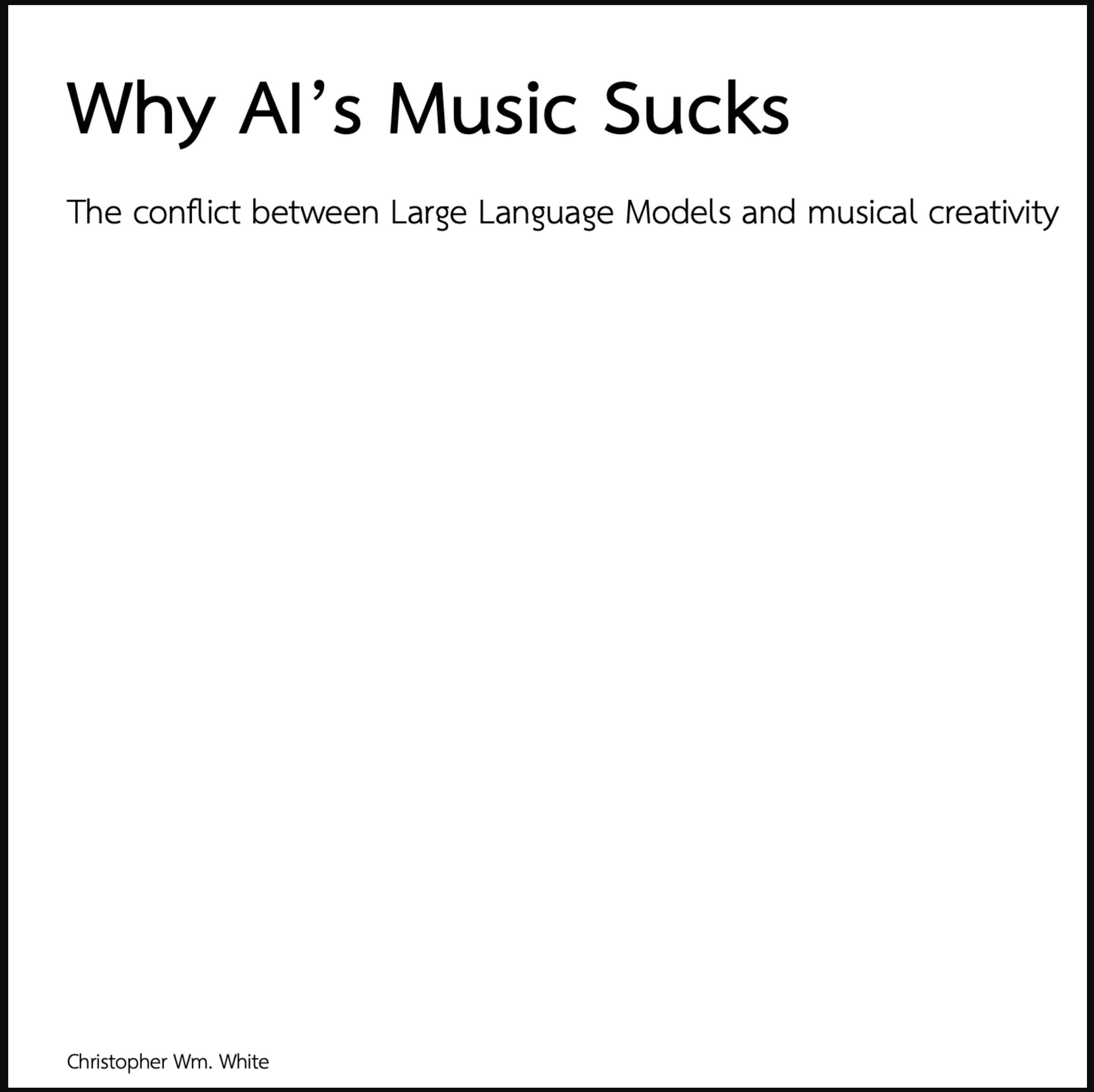
- There's little motivation behind musical AI, partly because we just don't want AI-made music
- This makes examples hard to come by, which makes it difficult to research how to represent music and to build models capable of understanding its structure
- Because of these dynamics, musical AI finds itself lagging behind AI in other media, with outputs that aren't particularly convincing to an audience

All of this is from my book draft,

and if you know any agents who might be interested, please let me know!

Book draft

cwmwhite@umass.edu


Why AI’s Music Sucks

The conflict between Large Language Models and musical creativity

Christopher Wm. White

Why AI’s Music Sucks

The conflict between Large Language Models and musical creativity

Christopher Wm. White

Thank you!

cwmwhite@umass.edu

Chris White
Dept. of Music and Dance
UMass Amherst

College of Humanities & Fine Arts