6. Symbolic vs audio generation

Generative Music AI

THE SOUND OF AI

upf. Universitat Pompeu Fabra Barcelona

MTG Music Technology Group

Overview

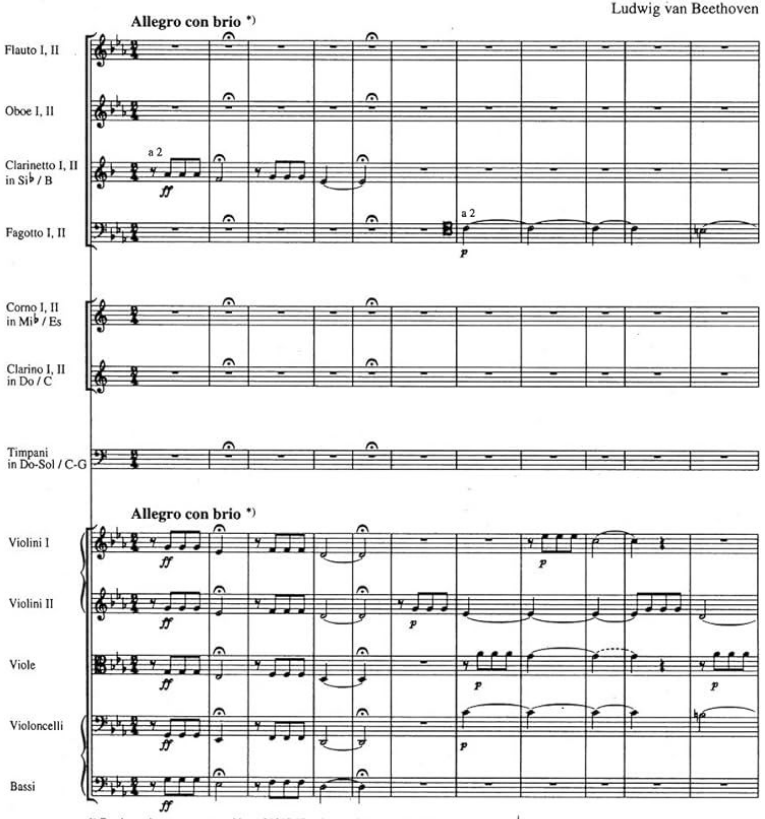
1. Music representation
2. Symbolic representation
3. Symbolic generation
4. Pros and cons of symbolic
5. Audio representation
6. Audio generation
7. Pros and cons of audio generation

À son Altesse Sérénissime
Monseigneur le Prince regnant de Lobkowitz
Duc de Raudnitz
et à son Excellence Monsieur le Comte de Rasumoffsky

Symphonie Nr. 5

c-moll
op. 67

Ludwig van Beethoven



*) Beethoven's metronome marking of 1817 / Beethovens Metronombezeichnung von 1817: ♩ = 108

Reconstruct *Another Brick in the Wall* from brainwaves

Music can be reconstructed from human auditory cortex activity using nonlinear decoding models (Bellier et al., 2023)

Encode music in a digestible format for a machine

Generative music system

A good music representation solves 50% of GM

Ideal music representation

- Objective and quantifiable
- Easy to manipulate
- Capture all musical details
- Compact


1P 35001 NIN 50000 2P 1
KO
99
KEN
Symbolic
Audio
ROUND 2
BATTLE 01


1P 35001 NIN 50000 2P 1
KO
99
Symbolic
Audio
ROUND 2
BATTLE 01


1P 35001 NIN 50000 2P 1
KO
99
Symbolic
Audio
ROUND 2
BATTLE 01


1P 35001
50000 2P 1
KO
99
Symbolic
Audio
ROUND 2
BATTLE 01


1P
50000
2P
KO
99
KEN
ROUND 2
BATTLE 01
Symbolic
Audio

Symbolic representation

- Symbols (e.g., notes, instruments)
- Similar to a score

Types of symbolic representations

Types of symbolic representations

- MIDI

Types of symbolic representations

- MIDI

(timestamp, midi note, velocity)

Types of symbolic representations

- MIDI
- MusicXML

```
<note>
  <pitch>
    <step>E</step>
    <alter>-1</alter>
    <octave>4</octave>
  </pitch>
  <duration>2</duration>
  <type>half</type>
</note>
```

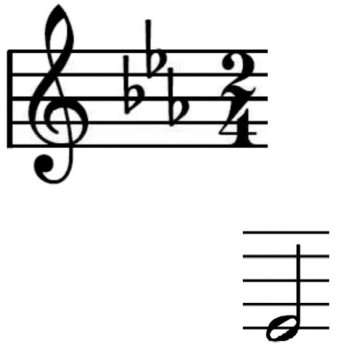
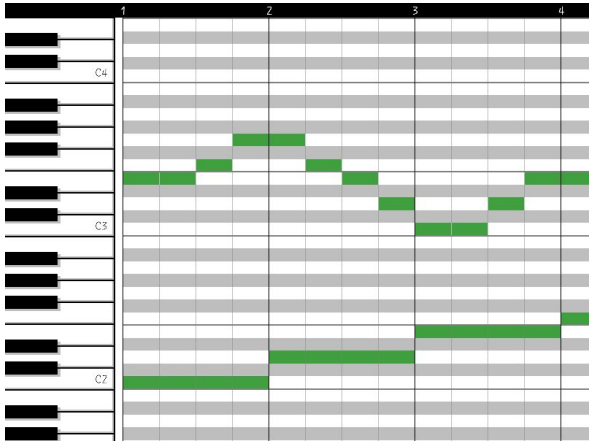


Figure 1.15 from [Müller, FMP, Springer 2015]

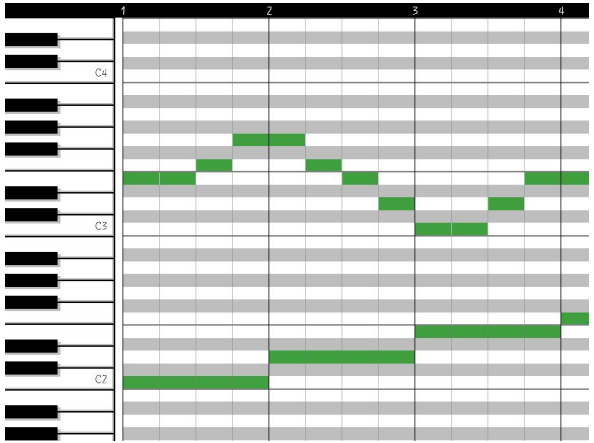
Types of symbolic representations

- MIDI
- MusicXML
- Piano-roll



Types of symbolic representations

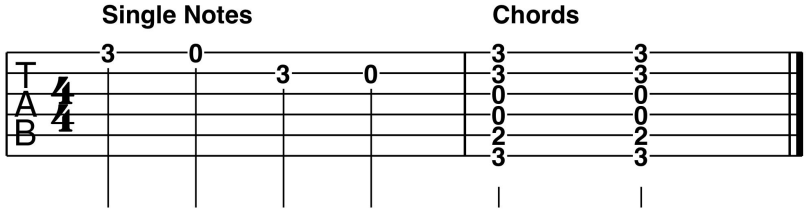
- MIDI
- MusicXML
- Piano-roll



$(pitch, t_{start}, t_{end})$

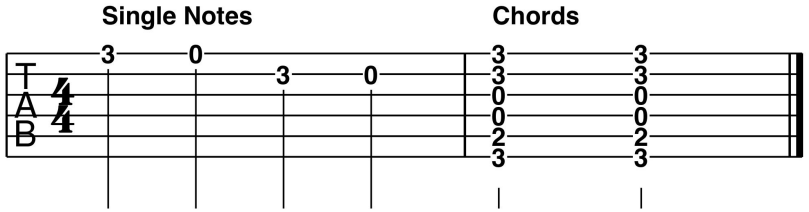
Types of symbolic representations

- MIDI
- MusicXML
- Piano-roll
- Tablature



Types of symbolic representations

- MIDI
- MusicXML
- Piano-roll
- Tablature



(string, fret, duration)

Types of symbolic representations

- MIDI
- MusicXML
- Piano-roll
- Tablature
- ABC notation

```
X:1
T:Twinkle, Twinkle, Little Star
M:4/4
K:C
C C G G | A A G2 | F F E E | D D C2 |
```

Types of symbolic representations

- MIDI
- MusicXML
- Piano-roll
- Tablature
- ABC notation
- Kern

```
*M4/4
*K[]
=1-
4c
4c
4g
4g
=2-
4a
4a
2g
=3-
4f
4f
4e
4e
=4-
4d
4d
2c
==
```

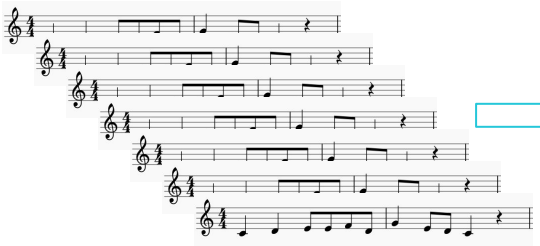
Types of symbolic representations

- MIDI
- MusicXML
- Piano-roll
- Tablature
- ABC notation
- Kern
- ...

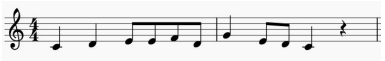
Symbolic: Discipline connections

- Music theory
- Composition
- (Computational) musicology

Symbolic generation



Generative music system



MuseNet (OpenAI, 2019)

- GPT2 architecture
- Trained on MIDI files
- Predict next token

Pros and cons of symbolic

- Compact
- Easy to manipulate
- Clear and precise
- Lots of compositional info
- Capture long-term dependencies
- Small models

Pros and cons of symbolic

- Compact
- Easy to manipulate
- Clear and precise
- Lots of compositional info
- Capture long-term dependencies
- Small models

- Oversimplified
- Musical limitations
- Limited performance info
- No production info
- Output isn’t audio

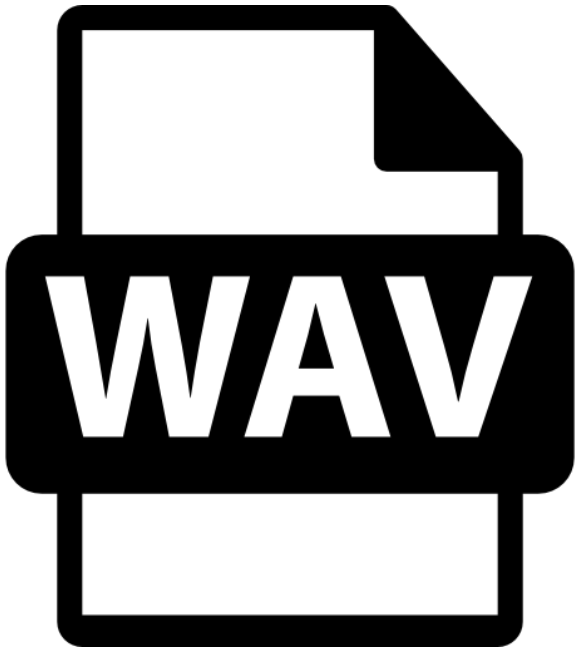
When is symbolic ideal?

- Structure + composition is focus
- Notated Western music (e.g., classical, jazz)

When isn’t symbolic ideal?

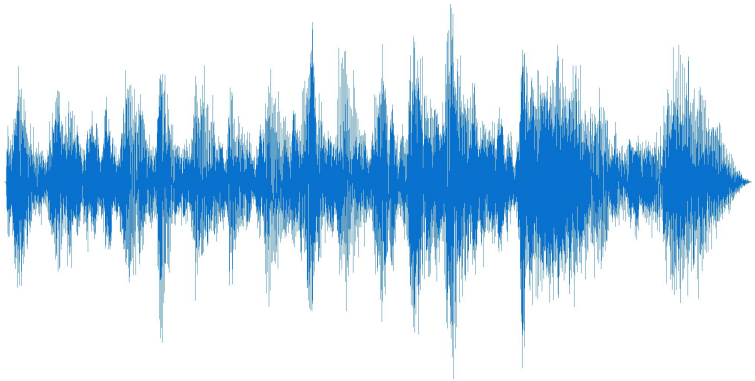
- Performance + production is focus
- EDM, drone, …

Audio representation



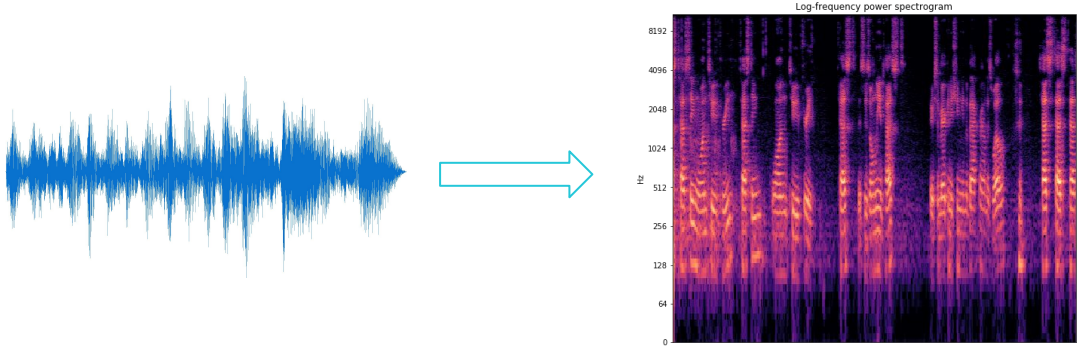
Types of audio representations

- Waveform



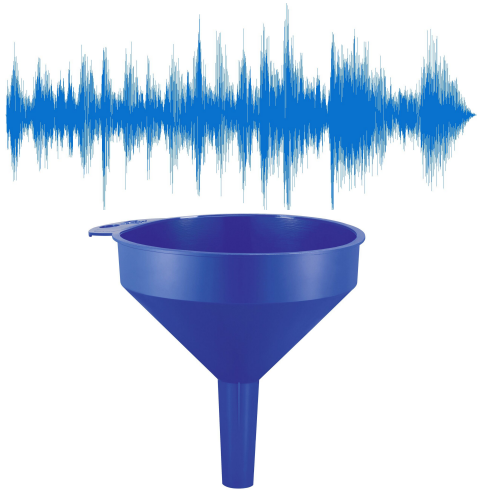
Types of audio representations

- Waveform
- Spectrogram



Types of audio representations

- Waveform
- Spectrogram
- Audio embeddings

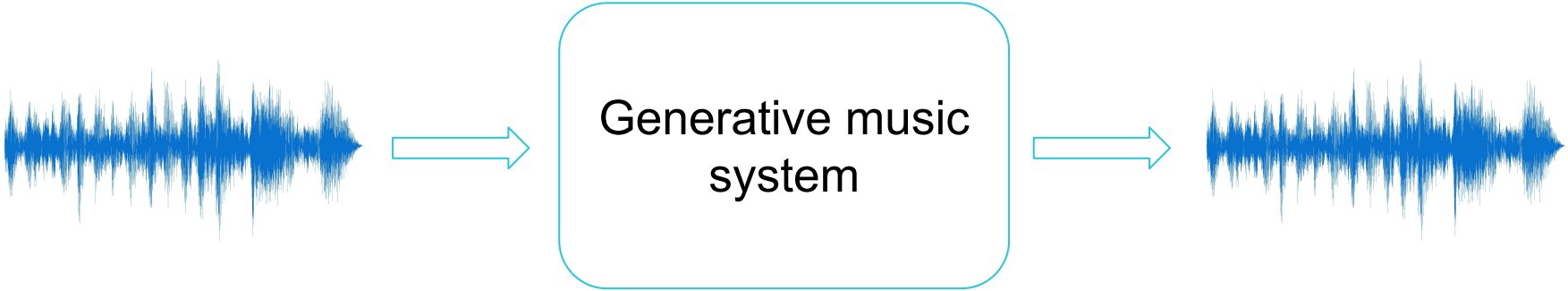


[0.34, 0.55, 0.23, 0.36, …]

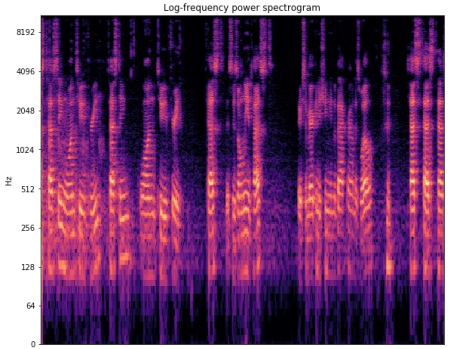
Audio: Discipline connections

- Digital signal processing
- Music information retrieval
- Sound design
- Music cognition

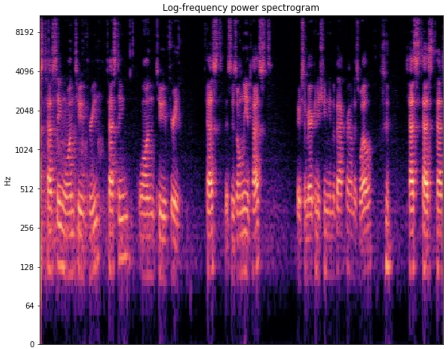
Audio generation: Waveform



Audio generation: Spectrogram



Generative music system



Sample audio-based models

- Jukebox (OpenAI)
- MusicLM (Google)
- MusicGen (Meta)
- RAVE (Ircam)

Pros and cons of audio generation

- Lots of performance info
- Lots of production info
- Complex and rich
- Audio output

Pros and cons of audio generation

- Lots of performance info
- Lots of production info
- Complex and rich
- Audio output

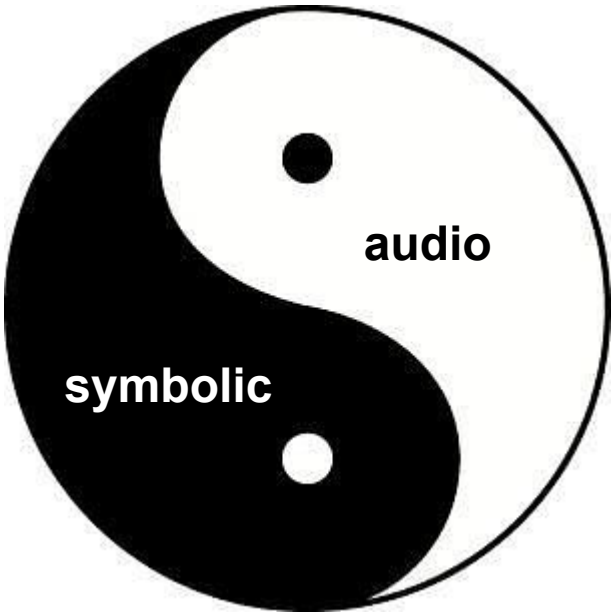
- Large dimensionality / size
- Difficult to manipulate
- No compositional info
- Model size
- Difficult to capture long-term dependencies

When is audio generation ideal?

- Performance + production is focus
- EDM, drone, …

When isn’t audio generation ideal?

- Structure + composition is focus
- Notated Western music (e.g., classical)


audio
symbolic

Key takeaways

- A good music representation solves 50% of GM

Key takeaways

- A good music representation solves 50% of GM
- Symbolic representation is similar to a score

Key takeaways

- A good music representation solves 50% of GM
- Symbolic representation is similar to a score
- Audio representation uses waveform or a transform

Key takeaways

- A good music representation solves 50% of GM
- Symbolic representation is similar to a score
- Audio representation uses waveform or a transform
- Symbolic has lots of compositional details

Key takeaways

- A good music representation solves 50% of GM
- Symbolic representation is similar to a score
- Audio representation uses waveform or a transform
- Symbolic has lots of compositional details
- Audio has lots of performance details

Key takeaways

- A good music representation solves 50% of GM
- Symbolic representation is similar to a score
- Audio representation uses waveform or a transform
- Symbolic has lots of compositional details
- Audio has lots of performance details
- Symbolic models are compact, but the output needs to be rendered to audio

Key takeaways

- A good music representation solves 50% of GM
- Symbolic representation is similar to a score
- Audio representation uses waveform or a transform
- Symbolic has lots of compositional details
- Audio has lots of performance details
- Symbolic models are compact, but the output needs to be rendered to audio
- Audio models are large, but directly generate audio output

What next?

Generative music techniques