

Generating Music with AI

# How to create controlled, high quality music

#1

# Music Generation with Language Model



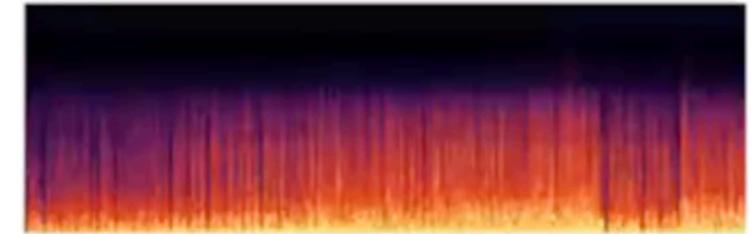
# Music Generation

- **Data representation**
  - Audio: waveform, spectrogram
- **Generation type**



waveform

- 1D vector
- High dimension (millions)
- Directly play sound

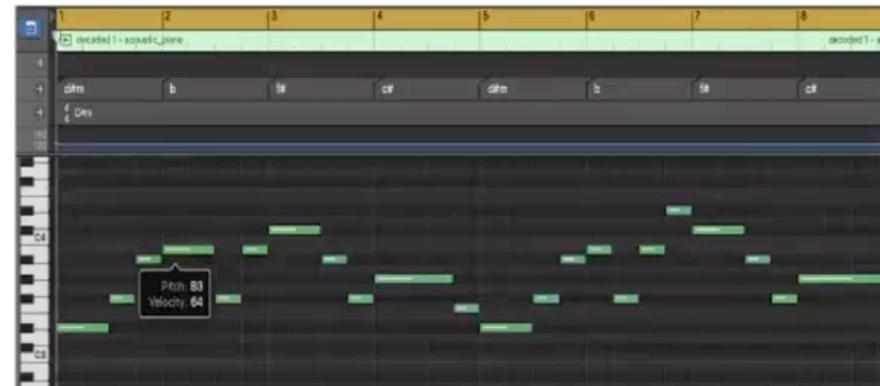


spectrogram

- 2D matrix
- Low dimension (tens of thousands)
- Convert to waveform to play sound

# Music Generation

- **Data representation**
  - Audio: waveform, spectrogram
  - **Symbolic: MIDI**
- **Generation type**



**MIDI** : Musical Instrument Digital Interface

- Standard protocol for making computer to play the score
- Can be represented as 1D vector with small dimension (hundreds)

# Music Generation

- Data representation
- **Generation type**

- **Continuation:** Music Transformer (Google Magenta), MuseNet (OpenAI)
- **Stylization:** MuseMorphose (Wu et al.)
- **Conditional generation**
  - **Style (genre, artist, text):** JukeBox (Dhariwal et al.), Riffusion (Forsgren et al.)
  - **Instrument:** Multi-track music machine (Ens et al.)

#2

**NeurIPS 2022, ComMU paper**

Create controlled, high quality music



# Intro.

- **Commercially viable music generation system**
  1. Controllability
  2. Diversity
  3. High-qualityFor the desired context

# Intro.

- **Commercially viable music generation system**
  1. Controllability
  2. Diversity
  3. High-qualityFor the **desired context** : composer's intension



Action scene



Jazz bar

# Intro.

- **Commercially viable music generation system**
  1. **Controllability** : can embrace desired context
  2. **Diversity** (composer's intention)
  3. **High-quality**For the **desired context**



Action scene



Jazz bar

# Intro.

- **Commercially viable music generation system**
  1. **Controllability** : “style, instrument, genre” → is it enough?
  2. Diversity
  3. High-qualityFor the desired context



Action scene



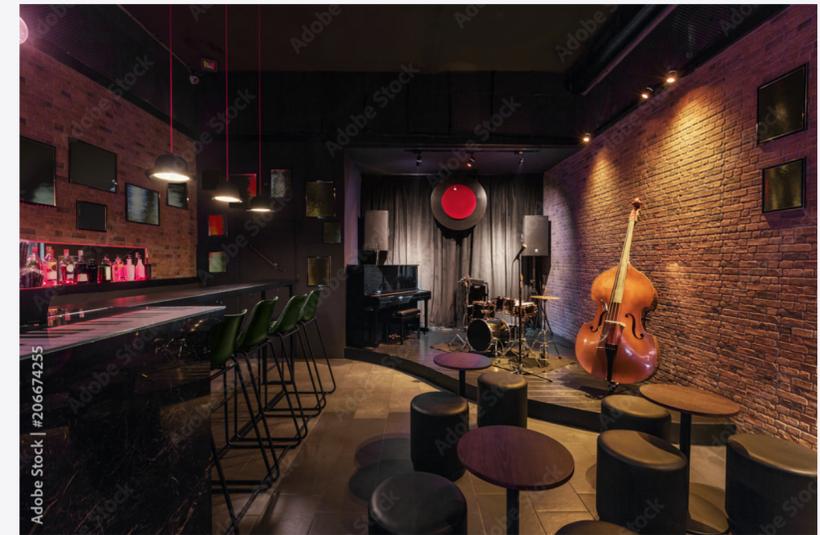
Jazz bar

# Intro.

- **Commercially viable music generation system**
  1. **Controllability** : 12 comprehensive & rich music metadata
  2. Diversity
  3. High-qualityFor the desired context



Action scene



Jazz bar

# Intro.

- **Commercially viable music generation system**
  1. Controllability
  2. Diversity
  3. High-qualityFor the desired context



Action scene



Jazz bar

# Intro.

- **Commercially viable music generation system**
  1. Controllability
  2. Diversity
  3. High-qualityFor the desired context



Action scene

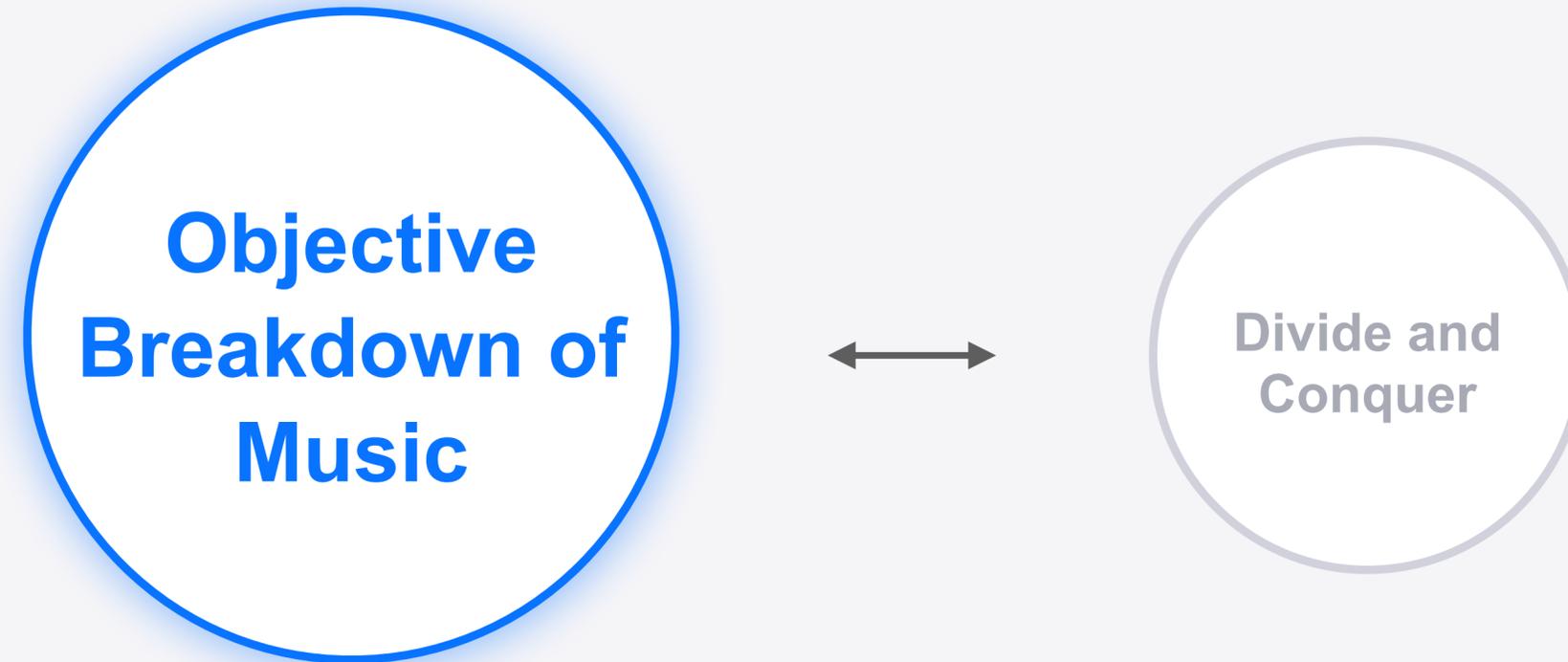


Jazz bar

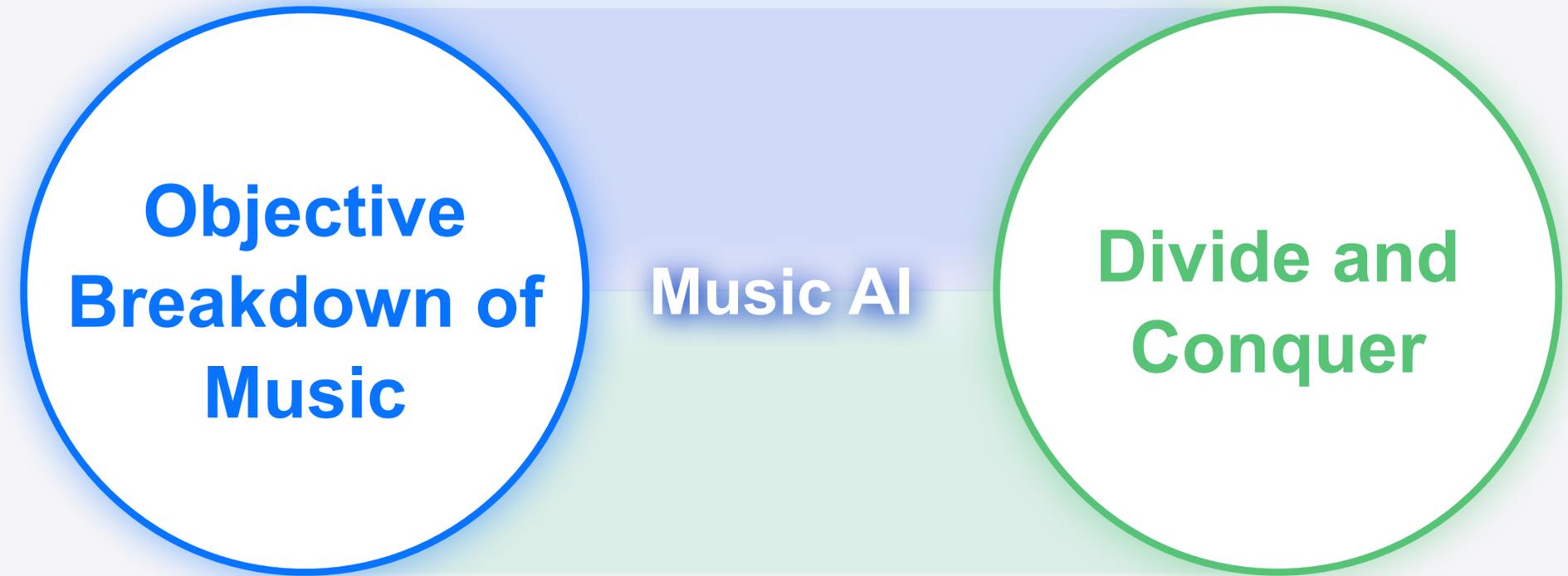
# Tackling Music Composition with AI



# Tackling Music Composition with AI



# Tackling Music Composition with AI



# Goal

- Generate **diverse, controllable, high-quality** music w/ **rich metadata** that embraces the **desired context**

## metadata

<b>bpm</b> : 100	<b>pitch</b> : mid_low
<b>genre</b> : cinematic	<b># of bar</b> : 8
<b>key</b> : C major	<b>chord</b> : F-C-Am-G
<b>inst.</b> : piano	<b>max vel</b> : 40
<b>track</b> : accompaniment	<b>min vel</b> : 36
<b>TS</b> : 4/4	<b>rhythm</b> : standard



## music sample

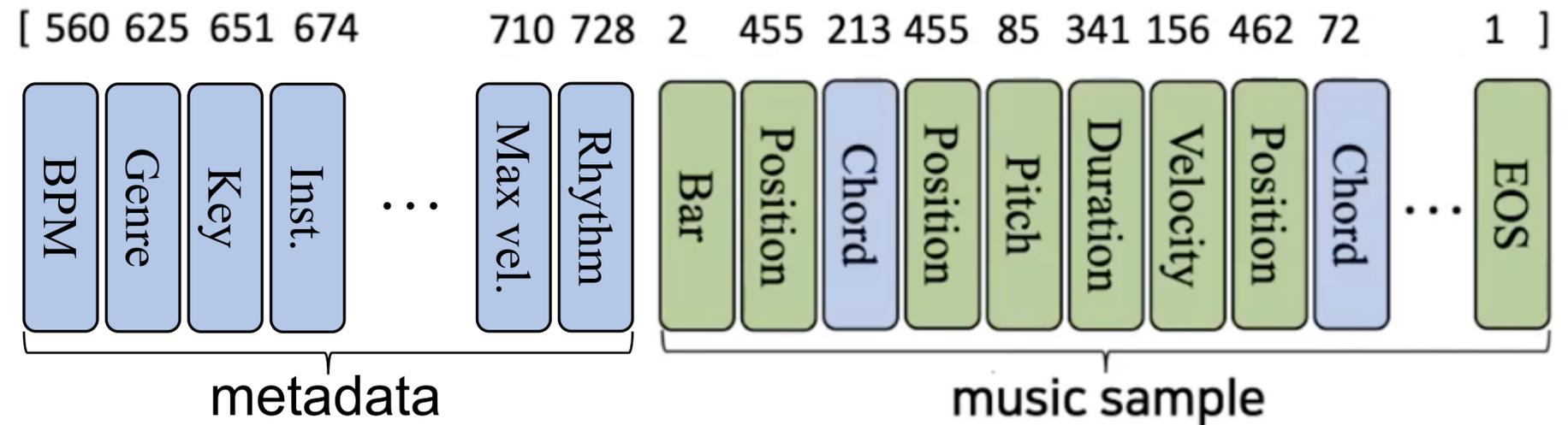


# ComMU representation

## metadata

<b>bpm</b> : 100	<b>pitch</b> : mid_low
<b>genre</b> : cinematic	<b># of bar</b> : 8
<b>key</b> : C major	<b>chord</b> : F-C-Am-G
<b>inst.</b> : piano	<b>max vel</b> : 40
<b>track</b> : accompaniment	<b>min vel</b> : 36
<b>TS</b> : 4/4	<b>rhythm</b> : standard

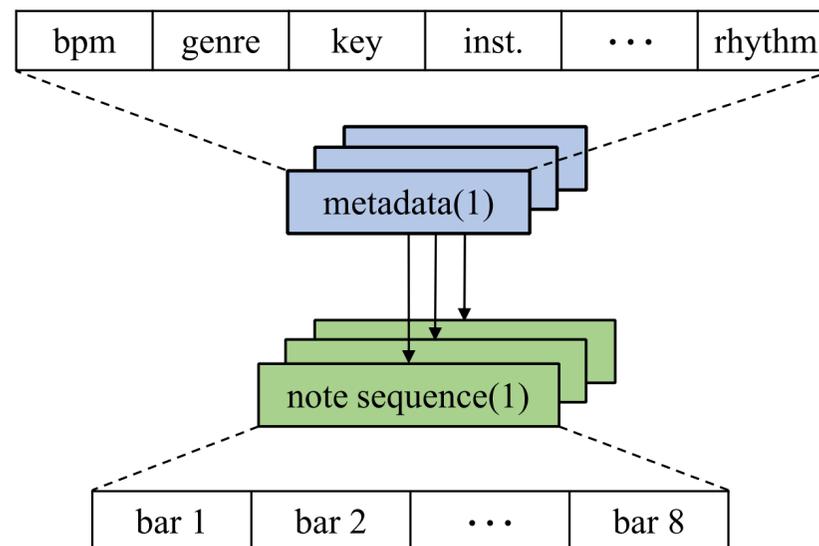
## music sample



# Combinatorial Music Generation

- New task to create diverse & high-quality music w/ given metadata

1. Create short note sequences w/ metadata



Stage 1

2. Combine sequences to produce a complete music

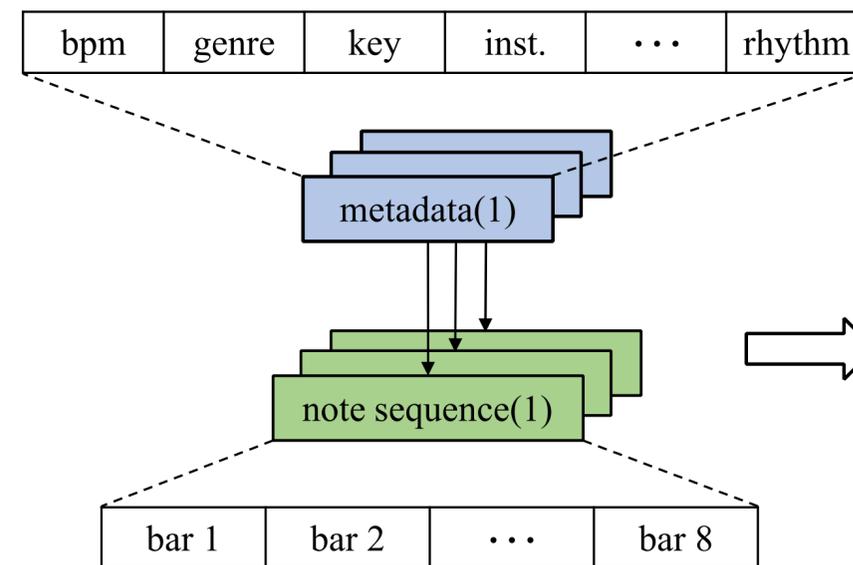
The diagram illustrates Stage 2. A blue arrow points from Stage 1 to a table. The table has five rows representing different musical parts: Main melody, Sub melody, Accompaniment, Bass, Riff, and Pad. The columns represent different note sequences. The 'Accompaniment' row is highlighted in green and contains 'note sequence(2)' in all three columns. The 'Main melody' row contains 'note sequence(1)' in the second column and 'note sequence(4)' in the third column. The 'Riff' row contains 'note sequence(3)' in the second and third columns. The 'Pad' row contains 'note sequence(5)' in the first column. Ellipses (...) are shown to the right of the table.

Main melody		note sequence(1)	note sequence(4)
Sub melody			
Accompaniment	note sequence(2)	note sequence(2)	note sequence(2)
Bass			
Riff		note sequence(3)	note sequence(3)
Pad	note sequence(5)		

Stage 2

# Combinatorial Music Generation

- Leveraging numerous combinations at stage 1 and stage 2 → **diversity**
- stage 1 (12 metadata) and stage 2 (combination strategy) → **controllability**
- Training dataset is manually created by professional composers → **high-quality**



Stage 1

Main melody		note sequence(1)	note sequence(4)
Sub melody			
Accompaniment	note sequence(2)	note sequence(2)	note sequence(2)
Bass			
Riff		note sequence(3)	note sequence(3)
Pad	note sequence(5)		

Stage 2

# Result: Stage 1

## common metadata

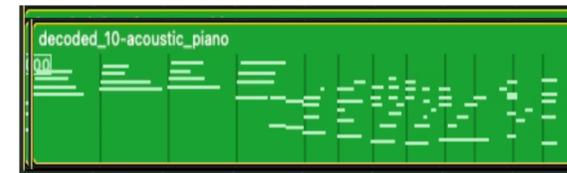
bpm : 80  
genre : cinematic  
key : A minor  
# of bar : 8  
rhythm : standard  
TS : 4/4

## individual metadata

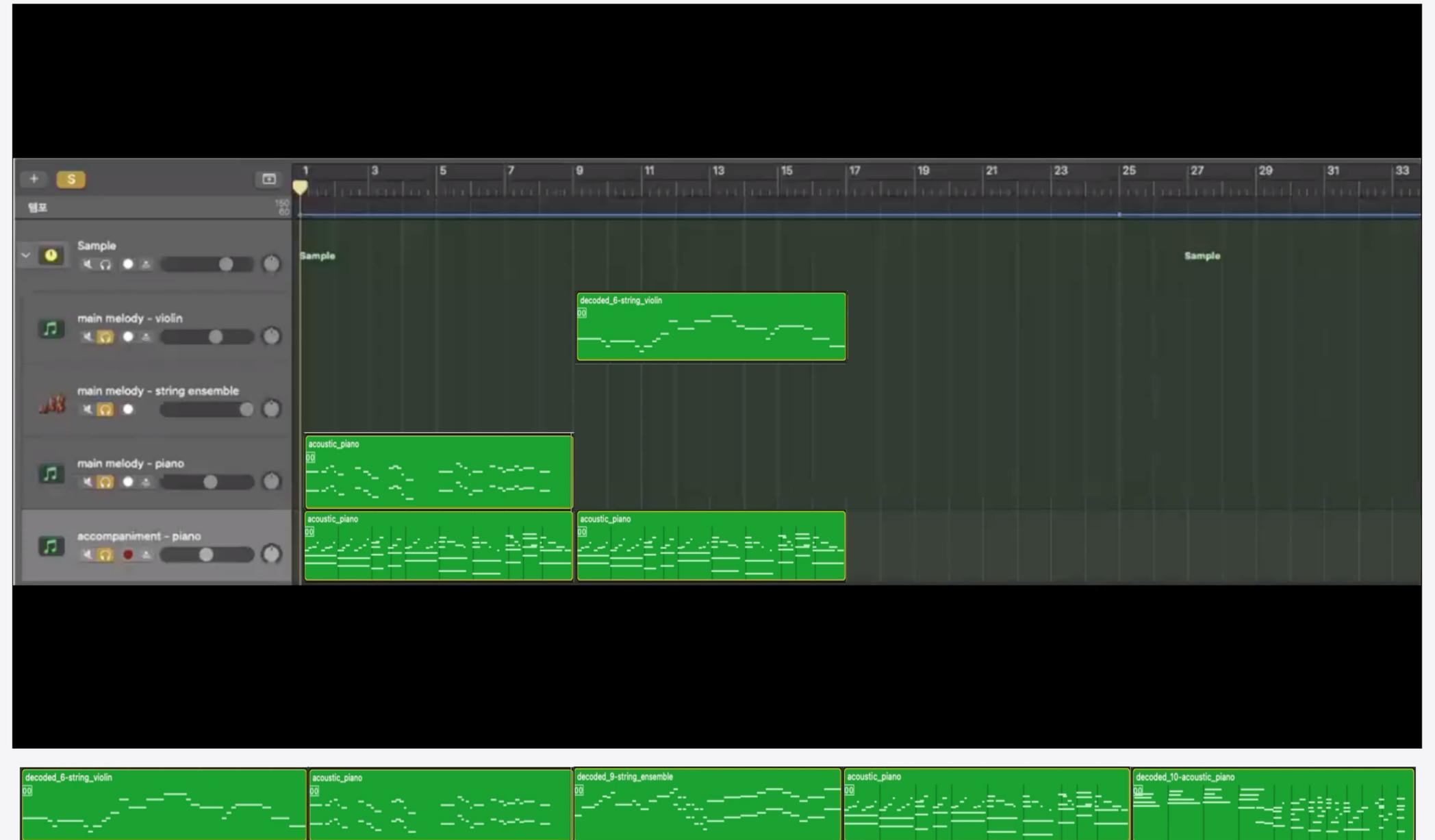
pitch : mid high  
track : accompaniment  
min/max vel : 25,100  
Inst. : piano  
Chord : Am7 - Gadd2 - F7 - A7 - Dm7 - E7 - F7 -  
Dm7 - Am7 - Bb7 - Gsus4 - Am7



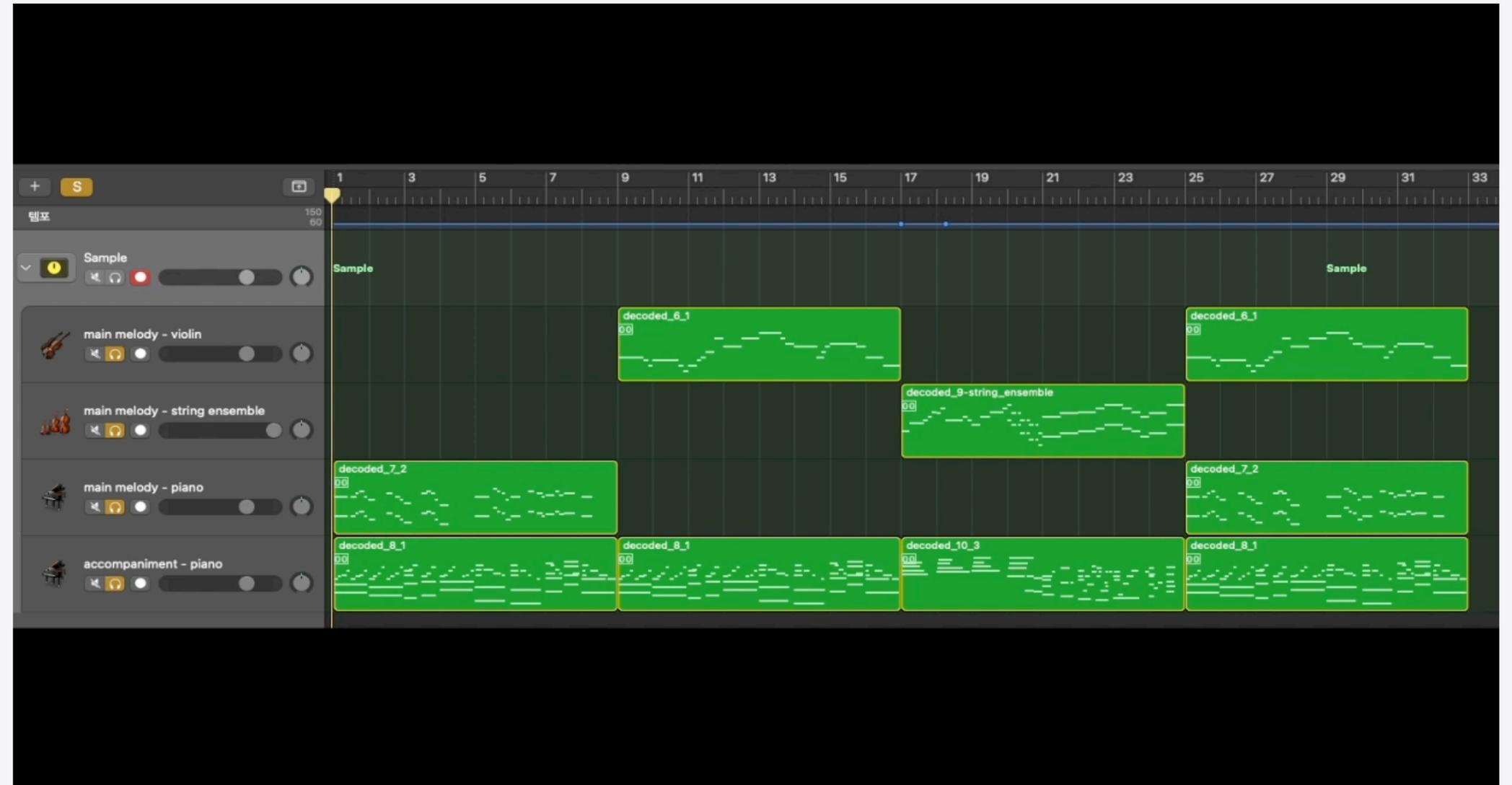
## note sequence



# Result: Stage 2



# Result: Stage 2



**Ignite your creativity.**