





Music GenAI Valerio #1 - Intro

<input checked="" type="checkbox"/> Favorite	<input type="checkbox"/>
<input checked="" type="checkbox"/> Archived	<input type="checkbox"/>
↗ Area/Resource	 Machine Learning
↗ Project	 GenAI IntroZ
☰ 🔍 Recipe Tags	

Course Goals

- Outline generative music **use cases**
- Review notable **symbolic- and audio-**based generative systems
- Use existing systems to **create music**
- **Develop** your own!
- Evaluate **limitations** of current systems
- **Analyse new approaches** to gen music
- Discuss **ethical**, regulatory aspects

Sign up to The Sound of AI **Slack** Community to join the discussion:

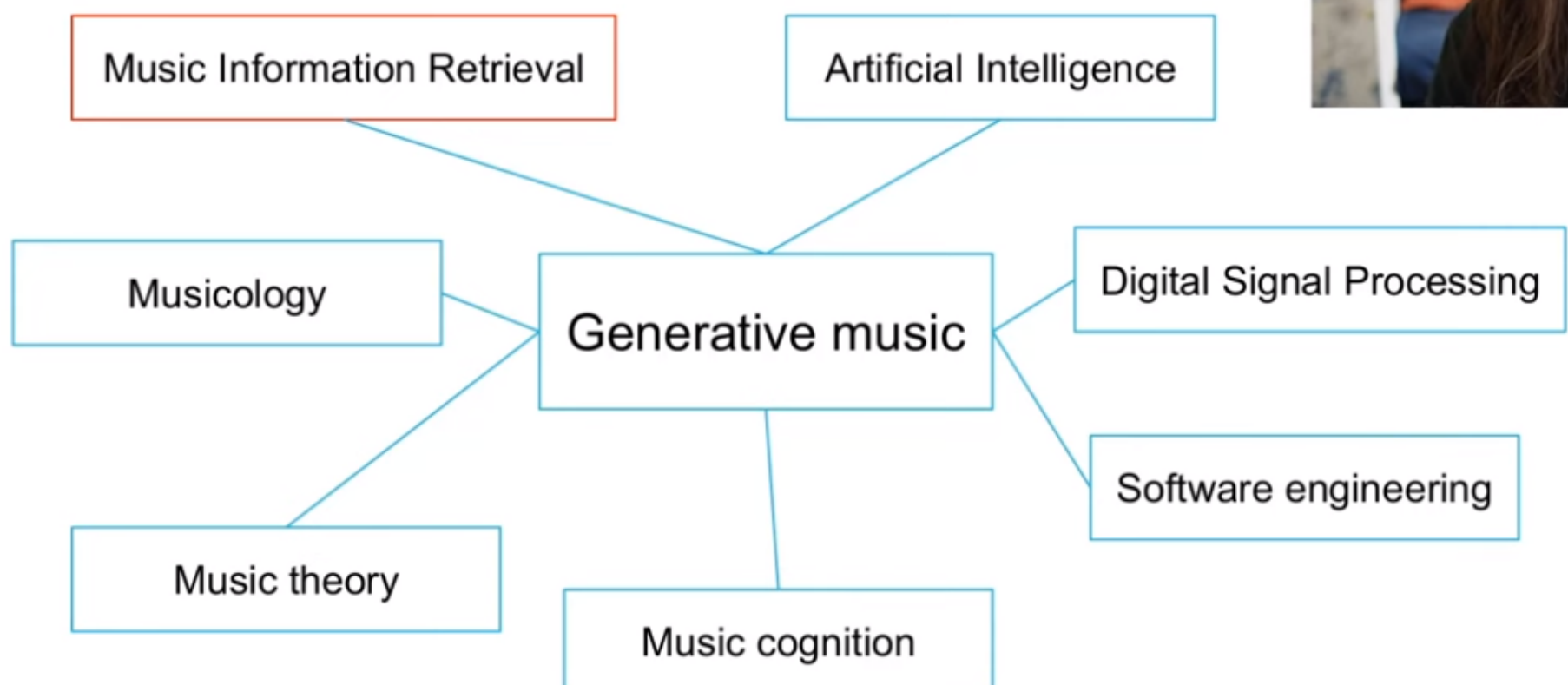
<https://valeriovelardo.com/the-sound-...>

Get the lecture **slides**:

<https://github.com/musikalkemist/gene...>

What's Generative Music?

The **art and science** of developing computer programmes that create music with a **varying degree of autonomy**.



GM Challenges

How can we represent music?

- **Formalizing the problem of music composition** is impossible due to its complexity and the multitude of factors and variables involved, ranging from the physics of sound to music perception and theory.
- Rules, spectrograms (audio), statistics, symbolic, embeddings - **good rep** is half the system - very “open” field

How can we evaluate the output of GM systems?

- **Intelligent tasks** can be reproduced with machines using various artificial intelligence techniques, but **creative tasks** are much more difficult to replicate due to the lack of a clear success metric and their ill-defined nature.
- Application specific

Who should evaluate GM systems?

- Not clear

History in 5 Eras with Notable Examples

Pre-computer era (1700 - 1956)

- Manual algorithms
- Composers lead the charge
- Randomness
- Re-combination
- Algorithms derive various musical parameters

Mozart Dice Game (1787)

- Random recombination
- 176 pre-composed bars
- Roll dice to select bars

Mode de valeurs et d'intensités (Messiaen, 1949)

- Parametrization of dynamics, articulation, pitch, duration
- Algorithm to select musical elements

Illiac Suite (Hiller & Isaacson, 1957)

- First computer-generated piece
- String quartet
- Four movements (different techniques → gen Grammers and/or markov chains)
- Different generative techniques

Academic era (1957 - 2009)

- Research activity
 - Lots of experimentation
 - Incremental advancements
 - Scattered community
- Musical output
 - Full-piece generation is rare
 - No focus on audio production quality
 - Score generation (symbolic)
 - Classical music

Experiments in Musical Intelligence (Cope, 1981)

- Invented because of composer's block
- Generate full piece
- Recombination approach
 - a. Analyse corpus
 - b. Extract signatures
 - c. Re-combine

First startup wave (2010 - 2016)

- Product focus
- Full-piece generation
- High-quality music datasets
- Machine learning



Melodrive (2016)

- Real-time video game music generation
- Music adapts to emotional context
- Unity SDK for indie game devs

Big tech experiments (2016 - 2022)

- Deep learning focus
- Massive datasets
- Massive computational power
- No commercial end goal



AWS DeepComposer (Amazon, 2019)



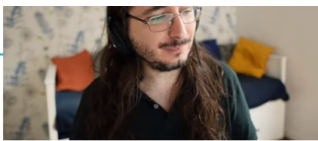
Jukebox (OpenAI, 2020)

- Raw-audio generation
- Advanced Deep Learning
- Full piece + lead vocals
- Performance details

Music AI hype (2023 - ?)

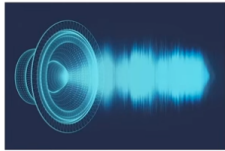
- Music industry is on fire
- Big tech explores commercial opportunity
- Scalable technology
- Really massive music datasets
- New startup wave

Text-to-music generation

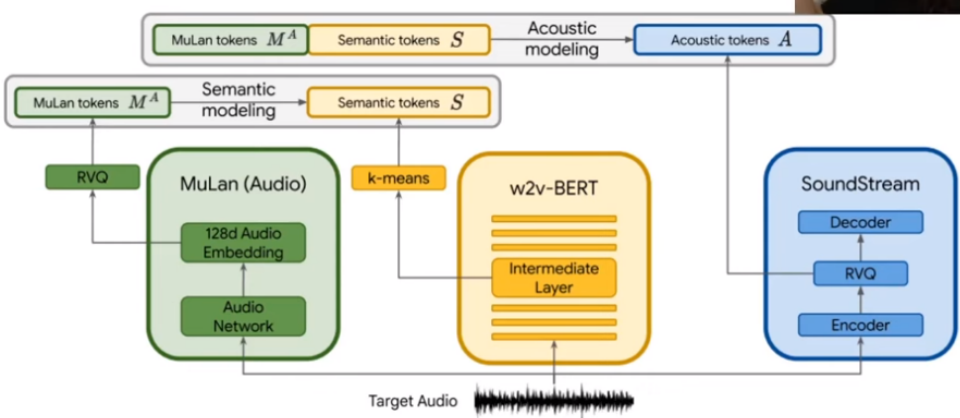


A fusion of reggaeton and electronic dance music, with a spacey, otherworldly sound.

Music model



MusicLM (Google, 2023)



MusicGen (Meta, 2023)

Simple and Controllable Music Generation

Jade Copet^{*}, Felix Kreuk^{*}, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve[†], Yossi Adi[‡], Alexandre Deffoes[‡]
^{*} equal contributions, [†] core team
Meta AI
{jadecopet, felixkreuk, deffoes}@meta.com

Abstract

We tackle the task of conditional music generation. We introduce MUSICGEN, a single Language Model (LM) that operates over several streams of compressed discrete music representation, i.e., tokens. Unlike prior work, MUSICGEN is comprised of a single-stage transformer LM together with efficient token interleaving patterns, which eliminates the need for cascading several models, e.g., hierarchically or upsampling. Following this approach, we demonstrate how MUSICGEN can generate high-quality samples, while being conditioned on textual description or melodic features, allowing better controls over the generated output. We conduct extensive empirical evaluation, considering both automatic and human studies, showing the proposed approach is superior to the evaluated baselines on a standard text-to-music benchmark. Through ablation studies, we shed light over the importance of each of the components comprising MUSICGEN. Music samples, code, and models are available at github.com/facebookresearch/audiocraft.

1 Introduction

Text-to-music is the task of generating musical pieces given text descriptions, e.g., “90s rock song with a guitar riff”. Generating music is a challenging task as it requires modeling long range sequences. Unlike speech, music requires the use of the full frequency spectrum [Müller, 2015]. That means sampling the signal at a higher rate, i.e., the standard sampling rates of music recordings are 44.1

Generative audio models launched in Jan ‘23

- Mousai
- AudioLDM
- SingSong
- RAVE 2
- Riffusion (Dec ‘22)
- ...

Second startup wave

