

Quick recap of text-to-audio (April 2024)

Roger Lam
lamroger.com

Lots of developments in 10 months

May 2023 - Google release MusicLM.

August 2023 - Meta release MusicGen.

Sept 2023 - Stability AI releases Stable Audio.

Dec 2023 - Suno releases Suno AI.

April 2023 - Stability AI releases Stable Audio 2.0.

Yesterday - Udio released as beta.

MusicLM is a decoder-only transformer model

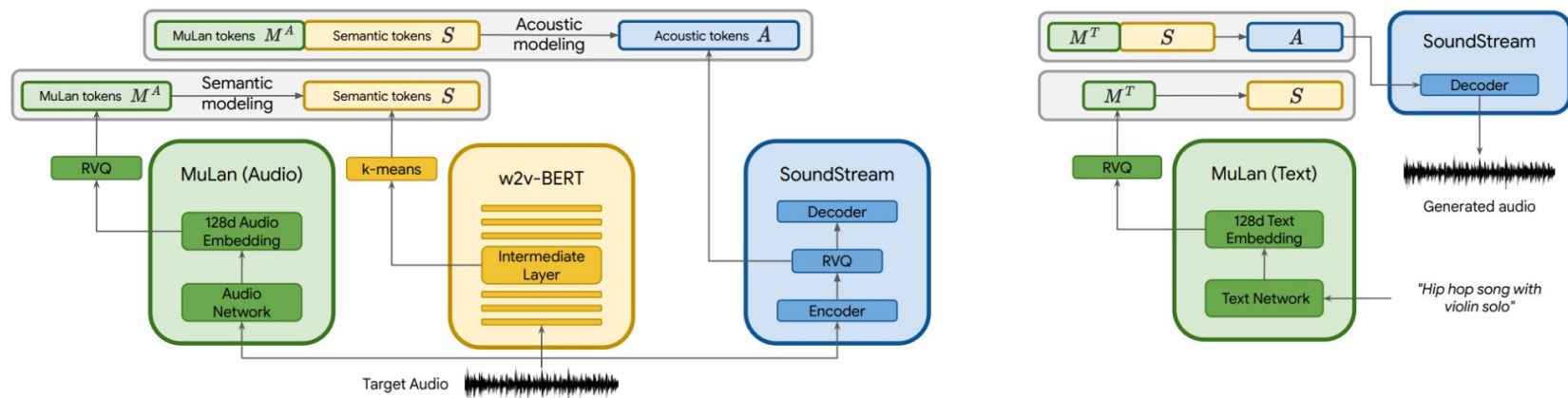


Figure 2. Left: During training we extract the MuLan audio tokens, semantic tokens, and acoustic tokens from the *audio-only* training set. In the semantic modeling stage, we predict semantic tokens using MuLan audio tokens as conditioning. In the subsequent acoustic modeling stage, we predict acoustic tokens, given both MuLan audio tokens and semantic tokens. Each stage is modeled as a sequence-to-sequence task using decoder-only Transformers. Right: During inference, we use MuLan text tokens computed from the text prompt as conditioning signal and convert the generated audio tokens to waveforms using the SoundStream decoder.

MusicGen uses a single end-to-end model

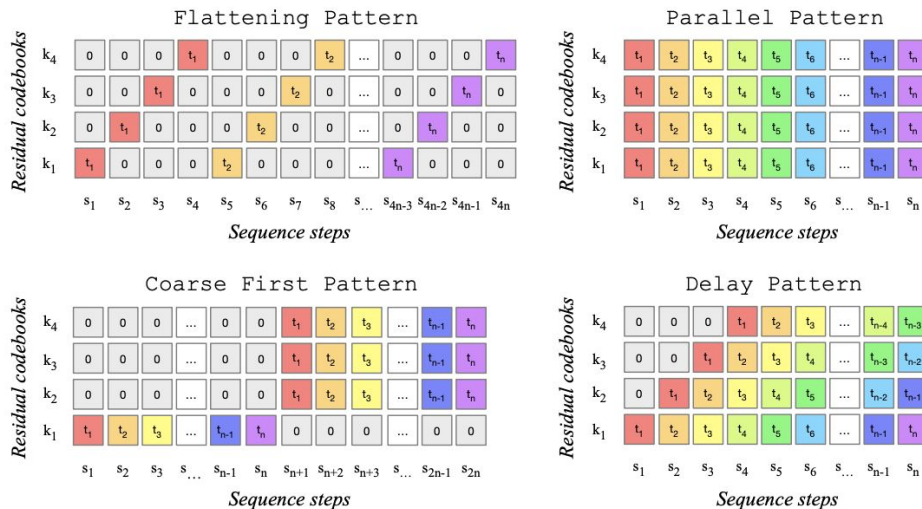
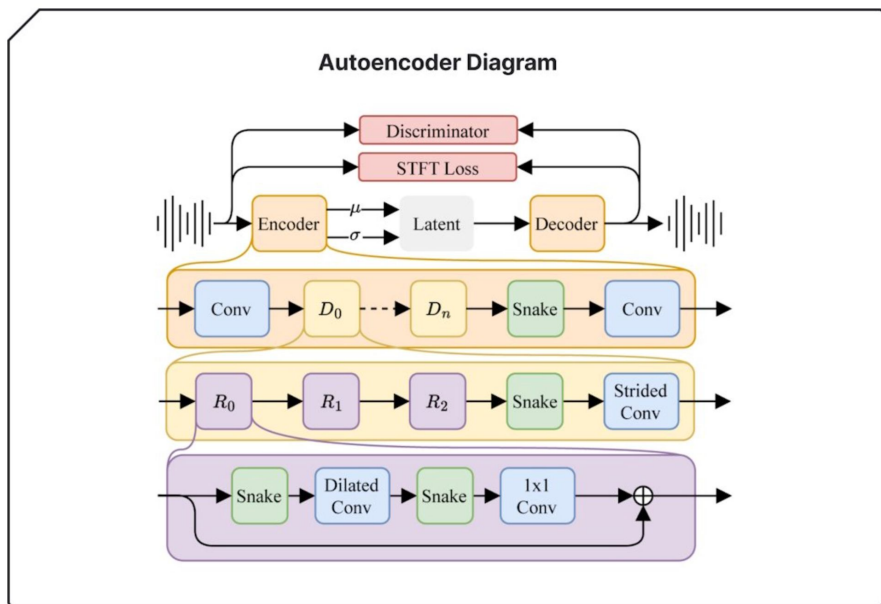
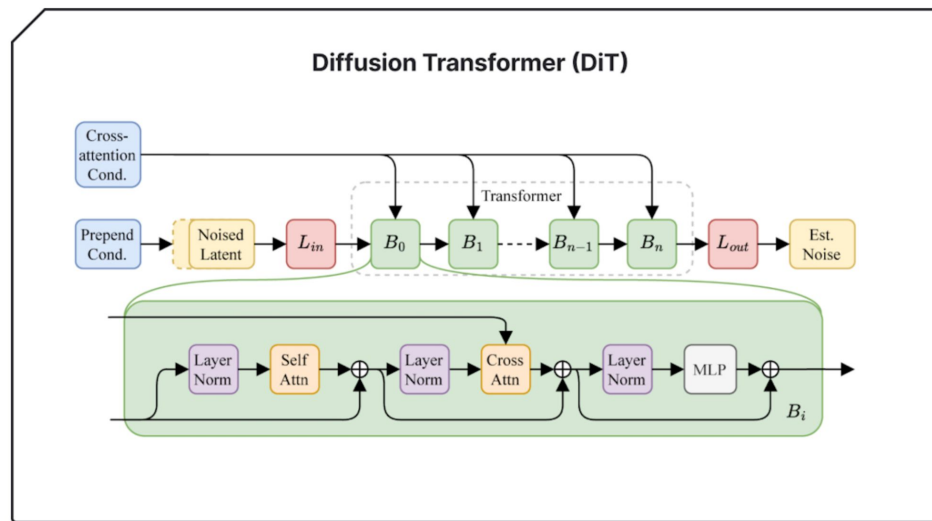


Figure 1: Codebook interleaving patterns presented in Section 2.2. Each time step t_1, t_2, \dots, t_n is composed of 4 quantized values (corresponding to k_1, \dots, k_4). When doing autoregressive modelling, we can flatten or interleave them in various ways, resulting in a new sequence with 4 parallel streams and steps s_1, s_2, \dots, s_m . The total number of sequence steps S depends on the pattern and original number of steps T . 0 is a special token indicating empty positions in the pattern.

Stable Audio 2 uses a Diffusion Transformer



The Autoencoder condenses audio and reconstructs it back to its original state. It captures and reproduces the essential features while filtering out less important details for more coherent generations.



A Diffusion Transformer (DiT) refines random noise into structured data incrementally, identifying intricate patterns and relationships. Combined with the Autoencoder, it gains the capability to process longer sequences to create a deeper, more accurate interpretation from inputs.

Suno is wild

Suno

Home
Create
Library
Explore **BETA**

Make a song about anything

An electronic future bass song about not being able to wait to see you again ↓


A groovy gospel song about being trapped in an AI song factory, help! ↓


An uplifting salsa song about the place where we used to go ↓


A dark indie song about a cozy rainy day Create 🎵


(You'll need to sign up for a free account)


Trending Now Show more +

#1  **Heavy Duty**
heavy metal, aggressive, trailer, clear sound, beat bass, epic drums, orchestra
▶ 27785 👍 589

#2  **I only ate 3 cheeseburgers!**
Country, catchy.
▶ 5981 👍 166

#3  **SPAGHETTI**
Italian bombastic theme with lots of horns and an energetic piano melody
▶ 182584 👍 2927

 **Hoppe, hoppe Reiter**
funny german rap
▶ 88200 👍 1627



SPAGHETTI

Italian bombastic theme with lots of horns and an energetic piano melody

April 9, 2024

👍 🗑️ 🔄 ⋮

Cook spaghetti in salted boiling water until al dente. Drain, reserving some pasta water.

In a skillet, heat olive oil over medium heat. Add chopped onion and cook until translucent, then add minced garlic and cook for a another minute.

Pour in canned diced tomatoes with their juices. Stir in dried oregano and basil. Season with salt and pepper. Optionally, add a pinch of sugar to balance acidity.

Simmer sauce for 15-20 minutes until slightl

Udio is at least equally wild

The screenshot shows the Udio beta website interface. At the top left is the 'udio beta' logo. A search bar contains the text 'a song about a moment of forgiveness' and a 'Create' button. The left sidebar has navigation options: 'Discover', 'My Creations', 'Liked Songs', and 'PLAYLISTS' with sub-options 'By you' and 'By others'. A large banner in the center reads 'Make your music' with a 'Learn how to create' button. Below this is a 'Staff Picks' section with six music recommendations, each with a unique cover image, title, artist, and play/like counts. The 'Trending' section at the bottom shows two more recommendations.

Staff Picks

- Echoes of Goodbye**
Conor
▶ 7350 ♥ 13
- Rising Sun Gospel**
charlie
▶ 9869 ♥ 31
- The Be Sharps - Baby On Board,...**
BobbyB
▶ 14744 ♥ 68
- Viscosity, Instrumental Hip Hop**
BobbyB
▶ 3432 ♥ 9
- Lyricless Thought**
Conor
▶ 3751 ♥ 11
- Neon Sax**
Conor
▶ 4223 ♥

Trending

- Allegro Consort in C**
Inever · 19h ago
▶ 9790 ♥ 52
- Wow..... I Didn't Know That [Full Track], Americana, Country**
BobbyB · 6d ago
▶ 24522 ♥ 117

Links

[MusicLM paper](#)

[MusicGen paper](#)

[Stable Audio 2](#)

[Suno](#)

[Udio](#)