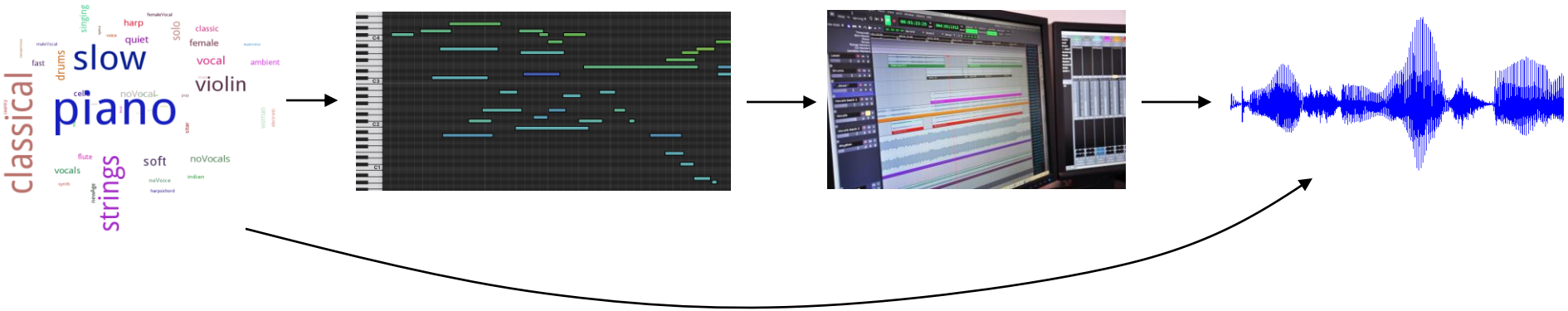
CTP431: Fundamentals of Computer Music

Overview of Music AI: Part 2

Juhan Nam

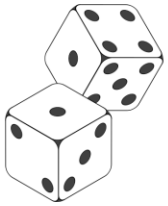
AI for Music Creation

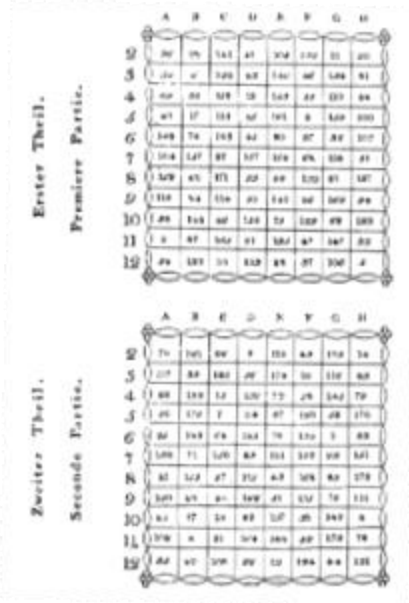
- Symbolic Music Generation
- Sound Synthesis and Digital Audio Effect
- Audio-Level Music Generation

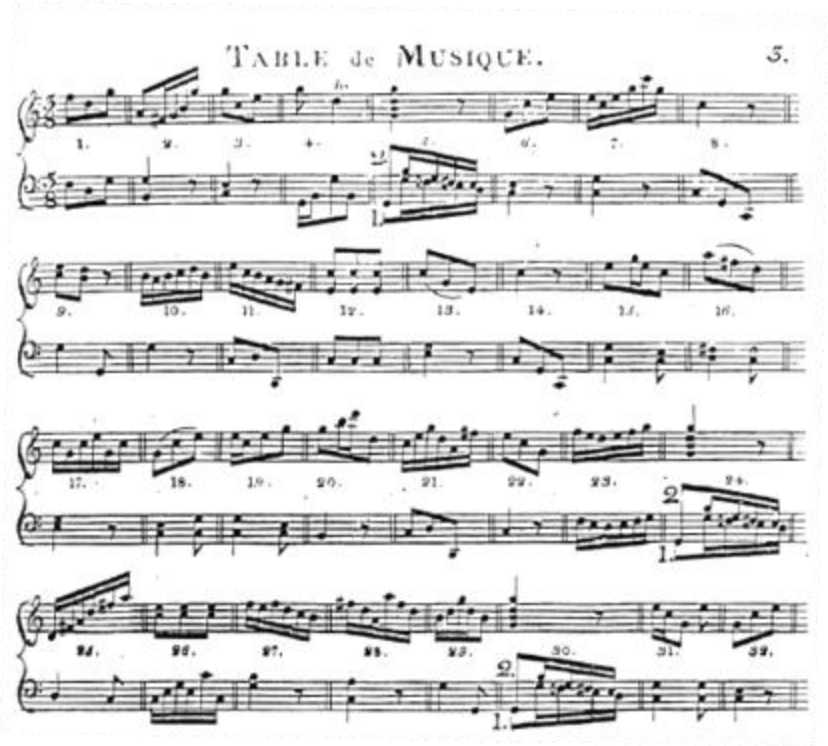


Musical Dice Game (1787)

- Mozart

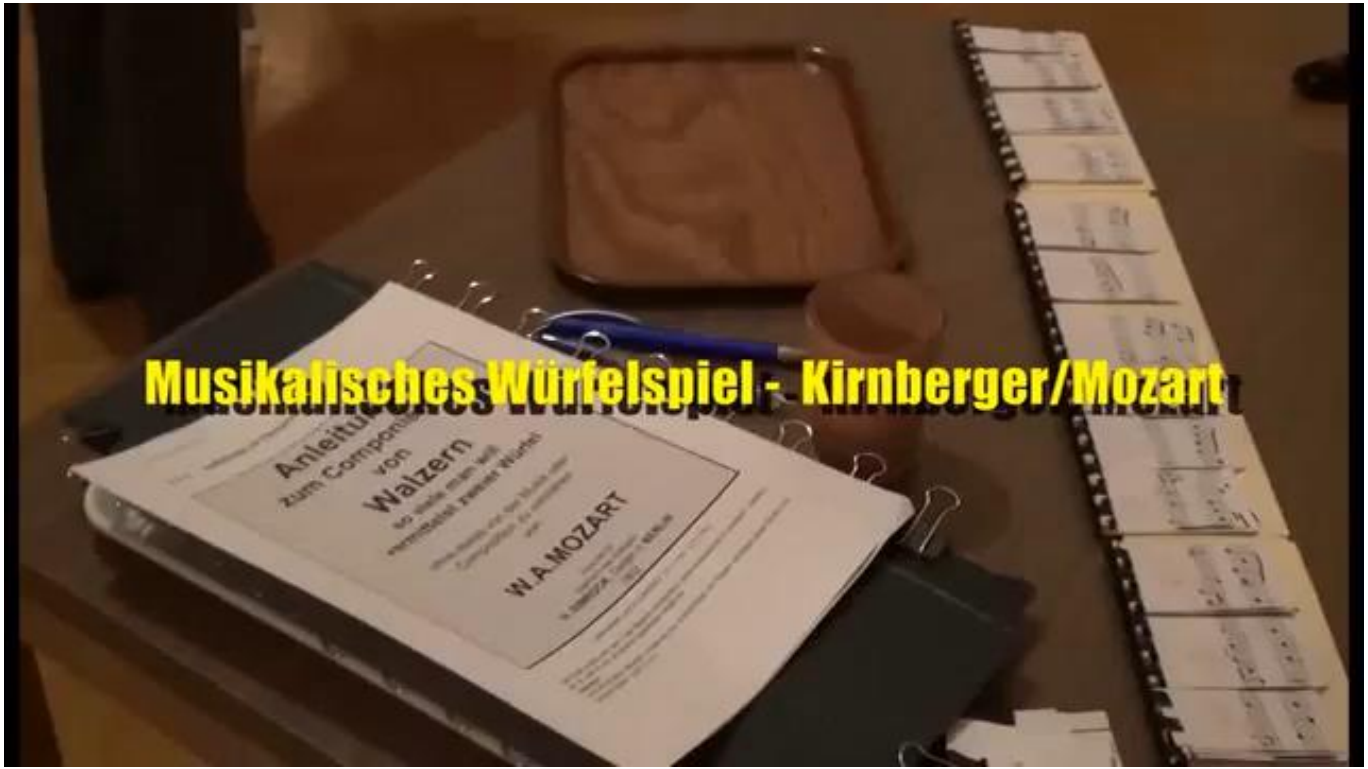






Mozart K. 516F

Musical Dice Game (1787)

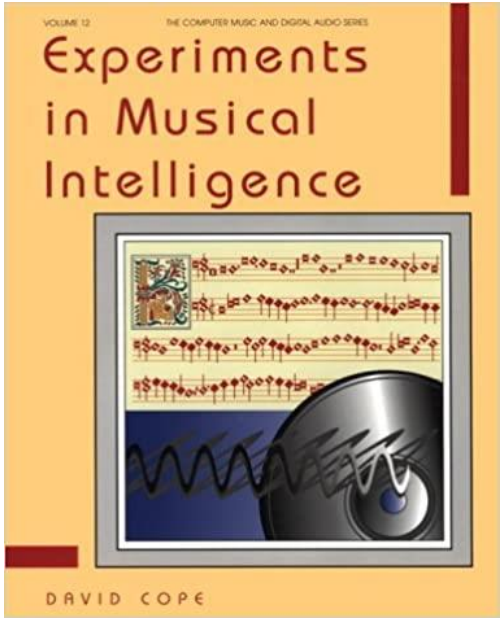


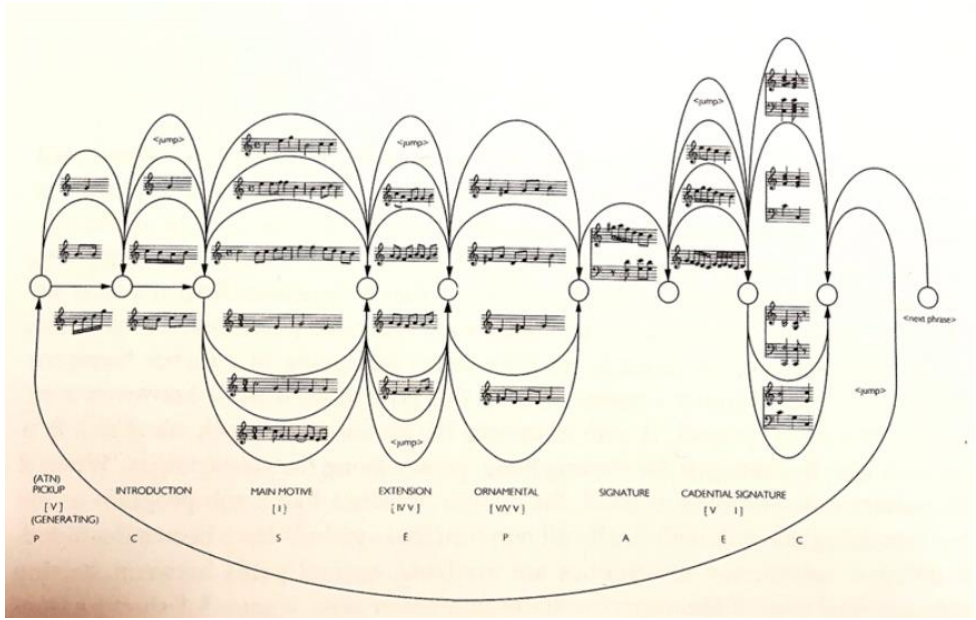
Illiac Suite (1956)

- Lejaren Hiller and Leonard Isaacson’s String Quartet
  - Markov model

Experiment in Music Intelligence (1980s)

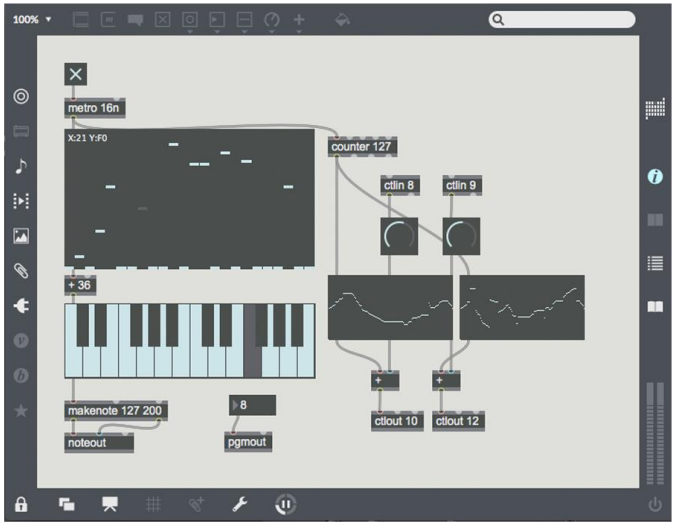
- David Cope
  - A rule-based natural language processing technique



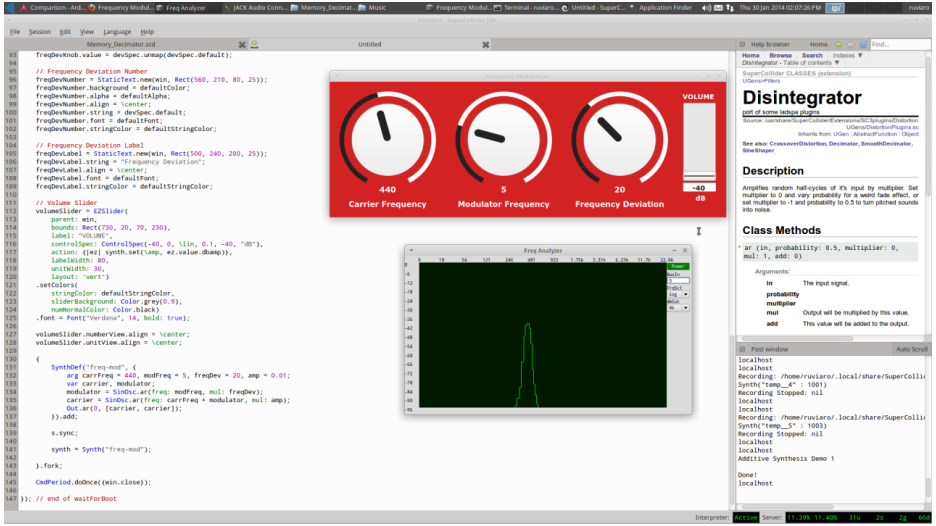


Augmented Transition Network

Algorithmic Composition



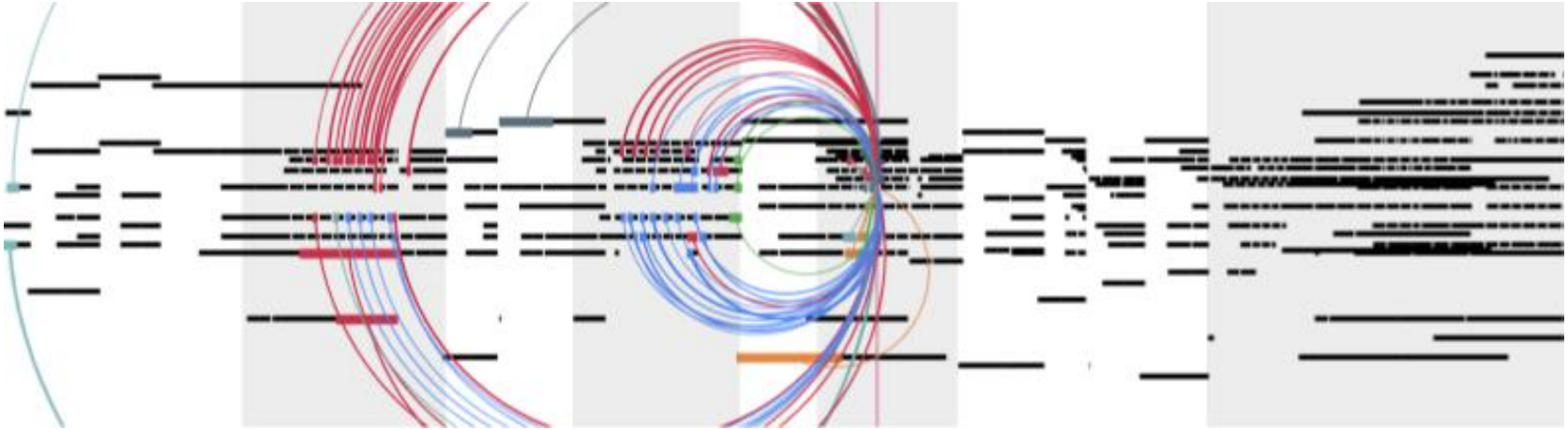
MAX



SuperCollider

Google Music Transformer (2018)

- Music Language Model



OpenAI MuseNet (2019)

MuseNet

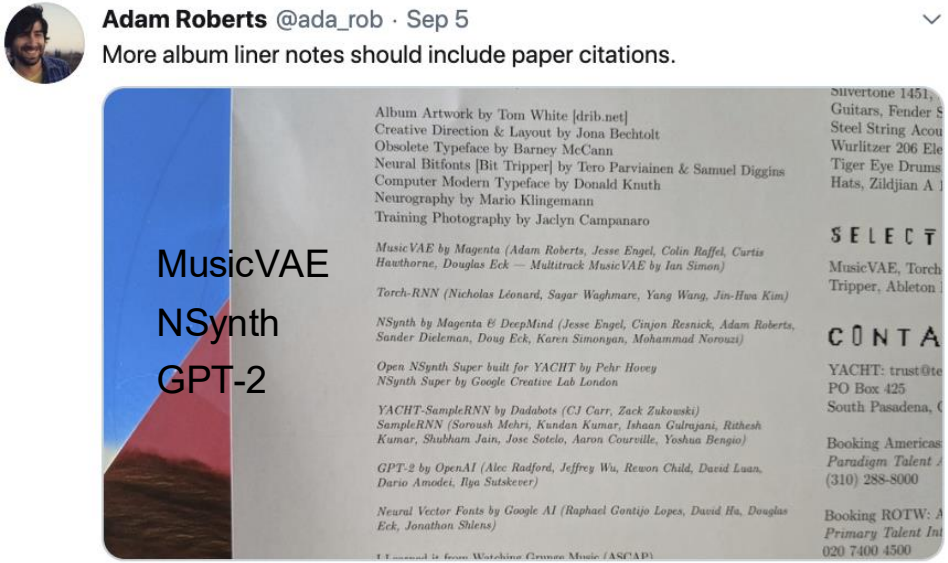
We've created MuseNet, a deep neural network that can generate 4-minute musical compositions with 10 different instruments, and can combine styles from country to Mozart to the Beatles.

MuseNet played an experimental concert on April 25th, 2019, livestreamed on OpenAI's Twitch channel, in which no human (including us) had heard the pieces before.

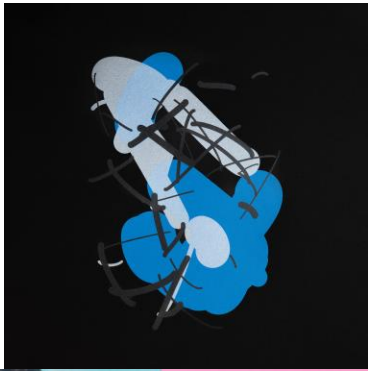
Magenta + YACHT (2019)

WHAT IS MAGENTA?

An open source research project exploring the role of machine learning as a tool in the creative process.



Chain Tripping



https://yacht.bandcamp.com/album/chain-tripping

Grammy-Nominated

AI Song Contest (2020 – 2022)

The AI Song Contest is an international competition exploring the use of AI in the songwriting process. Teams consisting of musicians, researchers, data scientists, developers and anyone else working on or interested in the combination of music and AI create a song using AI as part of their songwriting process.

2021 organizing team

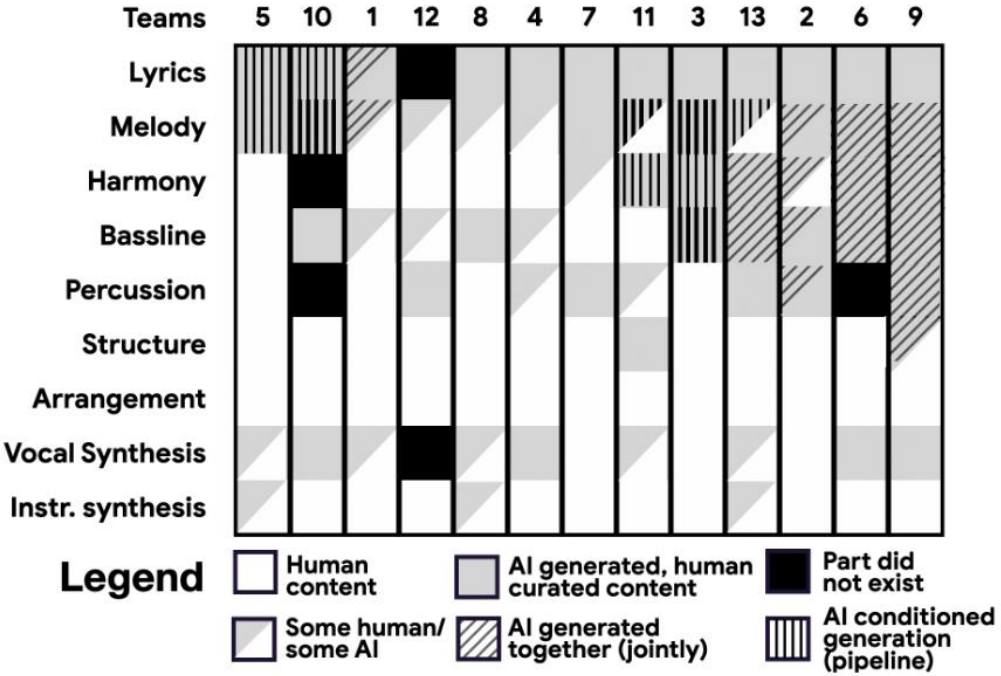
The 2021 contest was organised by Karen van Dijk, Rebecca Leger, Anna Huang, Ryan Groves, Rujing “Stacy” Huang, Hendrik Vincent Koops, Ashley Burgoyne, and Carol E. Reiley.

AI Song Contest 2020

Music building blocks	Models & techniques
Lyrics	GPT2, LSTM, Transformer
Melody	CharRNN, SampleRNN, LSTM + CNN, WaveNet + LSTM, GAN, Markov model
Harmony	LSTM, RNN autoencoder, GAN, Markov model
Bassline	LSTM + CNN, WaveNet + LSTM, GAN
Drums	DrumRNN, Neural Drum Machine, SampleRNN, Markov model
Multi-part	MusicVAE trio (melody, bass, drums), MiniVAE trio, Coconet/Coucou (4-part counterpoint), MusicAutobot (melody, accompaniment), Transformer (full arrangement)
Structure	Markov model
Vocal synthesis	WaveNet, SampleRNN, Vocaloid, Sinsy, Mellotron, Emvoice, Vocaloid, custom vocal assistant
Instrument synthesis	SampleRNN, WaveGAN, DDSP

“AI Song Contest: Human-AI Co-Creation in Songwriting”, Cheng-Zhi Anna Huang et al., ISMIR 2020

AI Song Contest (2020-2022)



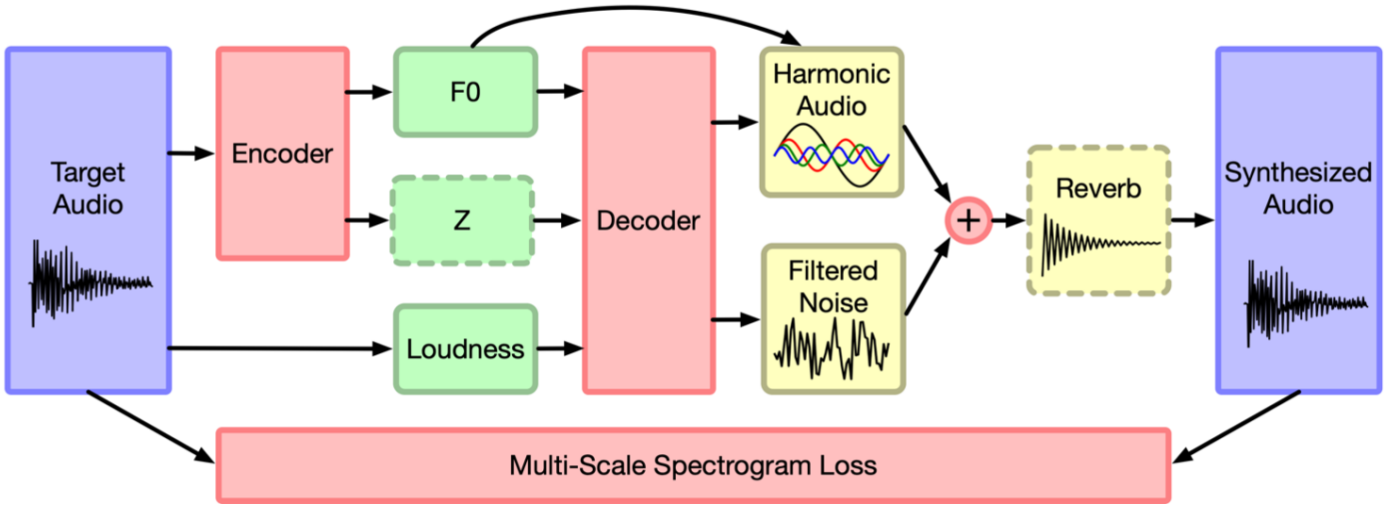
"AI Song Contest: Human-AI Co-Creation in Songwriting", Cheng-Zhi Anna Huang et al., ISMIR 2020

Tone Transfer

- Change timbre of the input source while preserving the expressions

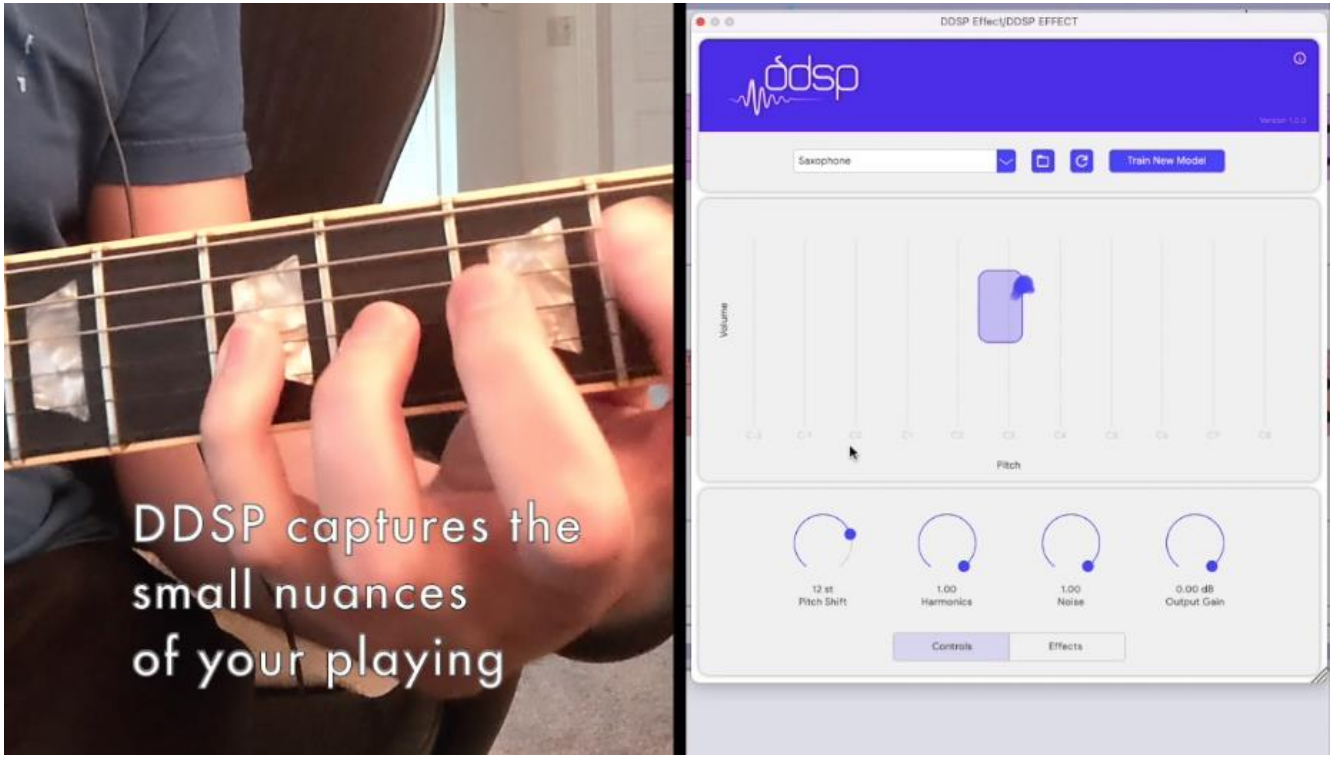
Differentiable Digital Signal Processing (DDSP)

- A hybrid model of neural network and spectral modeling synthesis



DDSP: Differentiable Digital Signal Processing, Jesse Engel, Lamtharn (Hanoi) Hantrakul, Chenjie Gu, Adam Roberts, ICLR, 2020

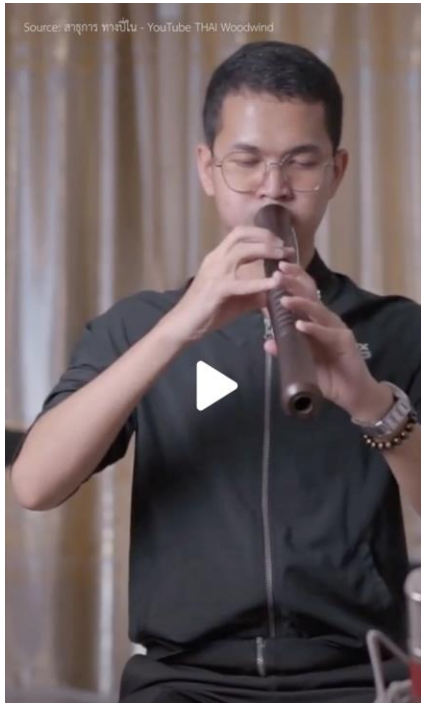
DDSP-VST

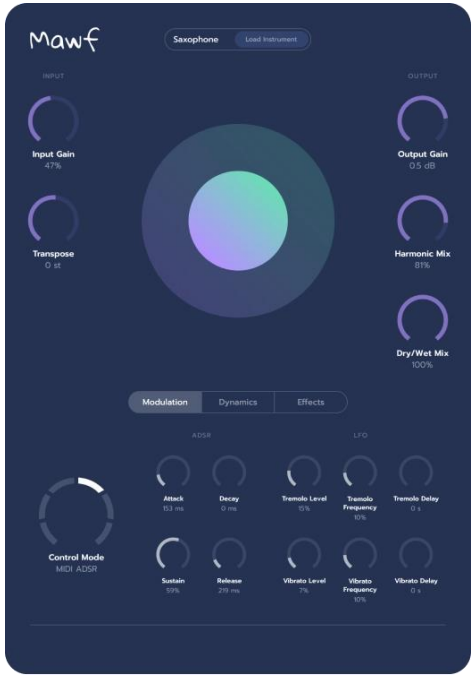


https://magenta.tensorflow.org/ddsp-vst

AI Song Contest 2022 Winner

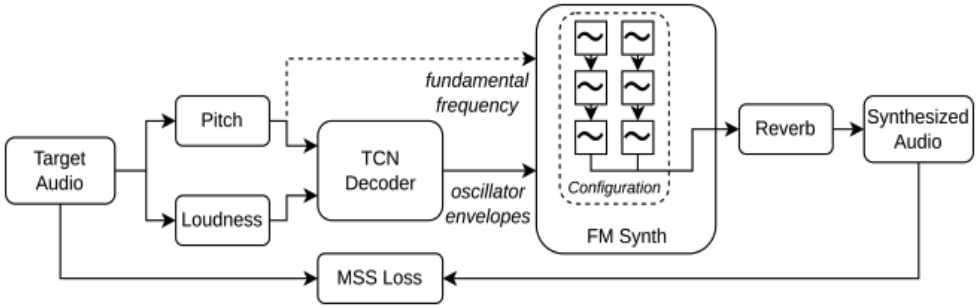
- Lamtharn “Hanoi” Hantrakul





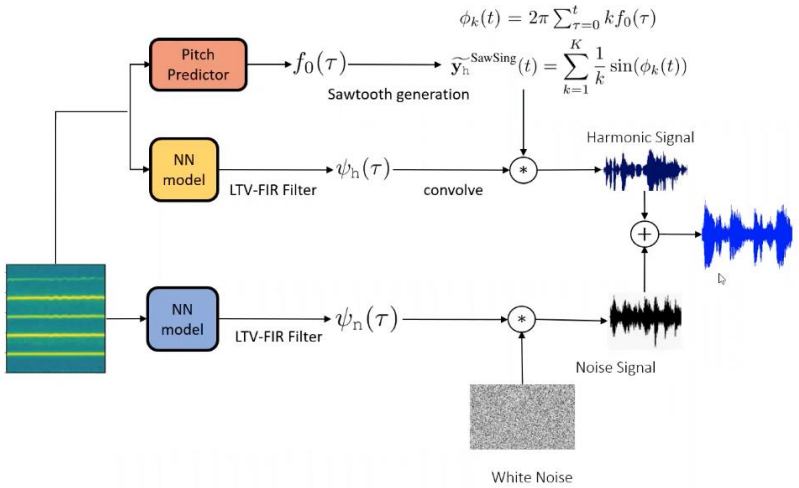
ByteDance Mawf

Neural Versions of Traditional Sound Synthesis Techniques



DDX7: Differential FM Synthesis

https://fcaspe.github.io/ddx7/



SawSing: Differential Subtractive Synthesis

https://ddspvocoder.github.io/ismir-demo/

Differentiable Wavetable Synthesis

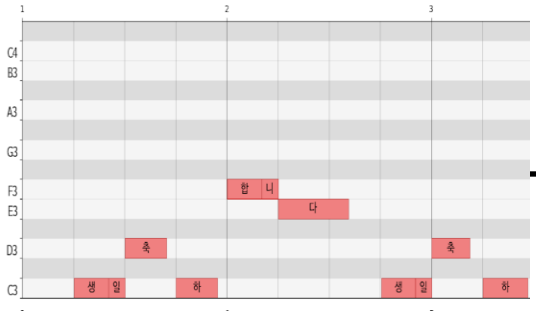
Neural Granular Sound Synthesis

DDS-Piano

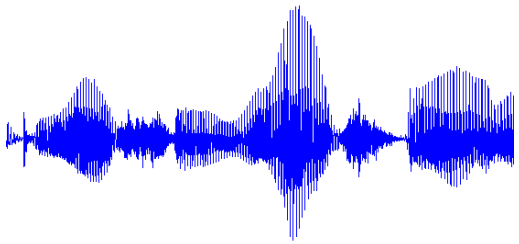
...

Singing Voice Synthesis (SVS)

- Generate singing voice audio from melody (MIDI) and lyrics (text)
  - Huge advances in neural SVS models over the past few years
  - Challenge: the lack of precisely annotated MIDI data to audio



Melody and Lyrics

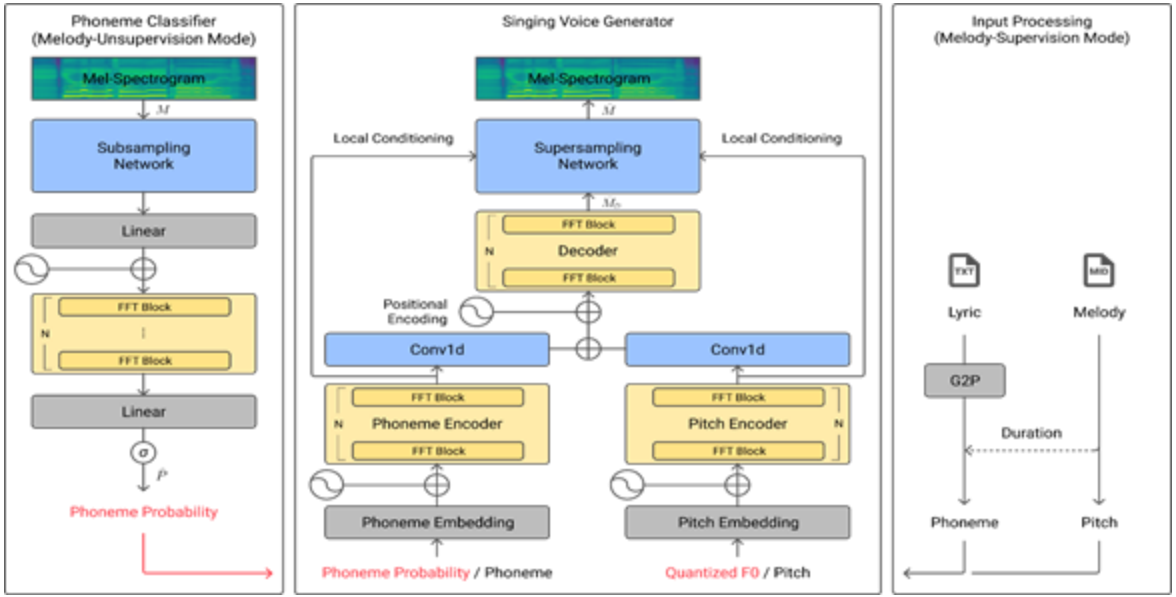


Audio

Singing Voice Synthesis

Singing Voice Synthesis Without MIDI Labels

- Train the model only with non-aligned text labels (like speech synthesis)
  - A phoneme classifier with the CTC loss: time-aligned phonemes (text input)
  - A pitch estimator: F0 (MIDI input)



A Melody-Unsupervision Model for Singing Voice Synthesis, Soonbeom Choi and Juhan Nam, ICASSP, 2022

Singing Voice Synthesis Services

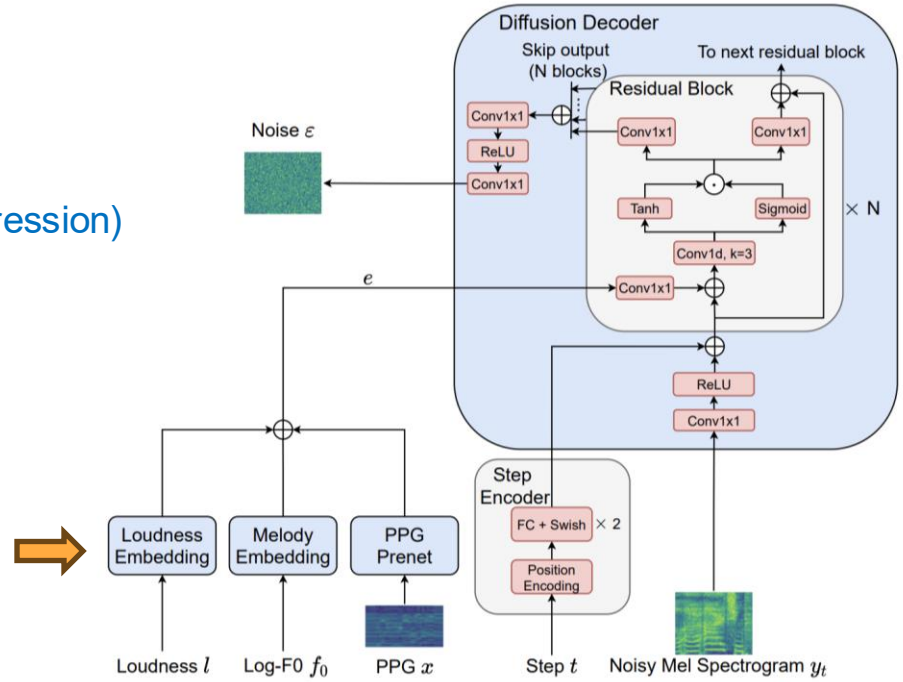
- Vox Factory: https://voxfactory.app/

AI Cover Song

- Singing Voice Conversion

Control Input
- Loudness, F0 (melody, expression)
- PPG (lyrics, phonetics)

(speaker-invariant)



Conditional Diffusion Model
- Timber

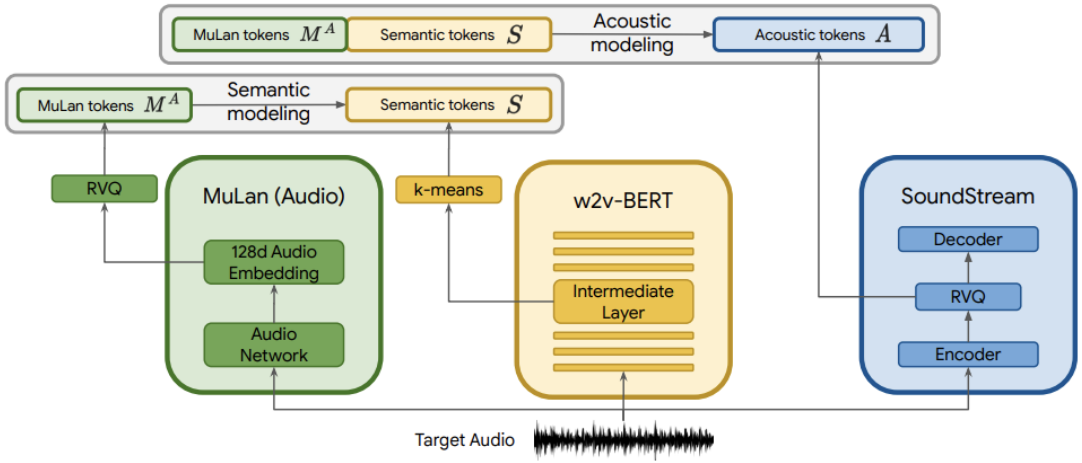
(speaker-specific)

https://www.youtube.com/watch?v=PmOpnSuGxbc

DiffSVC: A Diffusion Probabilistic Model for Singing Voice Conversion, Songxiang Liu, Yuewen Cao, Dan Su, Helen Meng, ARSU, 2021

MusicLM

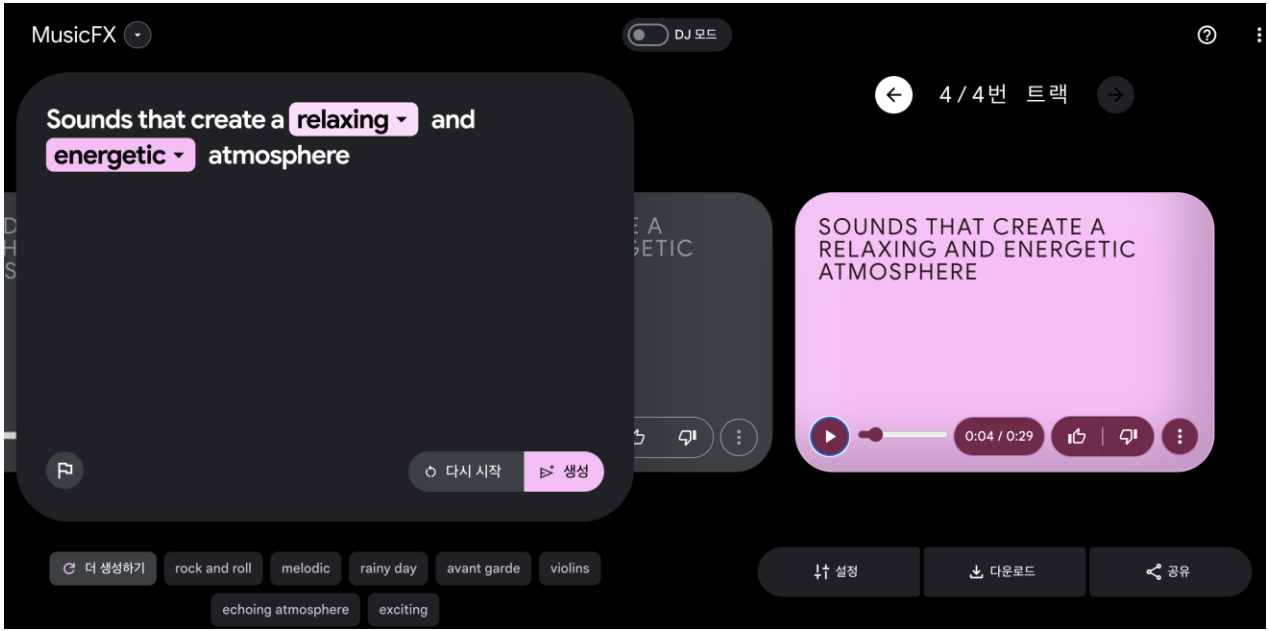
- Audio language model using acoustics tokens
  - text conditioning and/or melodic conditioning



https://google-research.github.io/seanet/musiclm/examples/

MusicLM: Generating Music From Text, Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, …, Arxiv, 2023

MusicLM

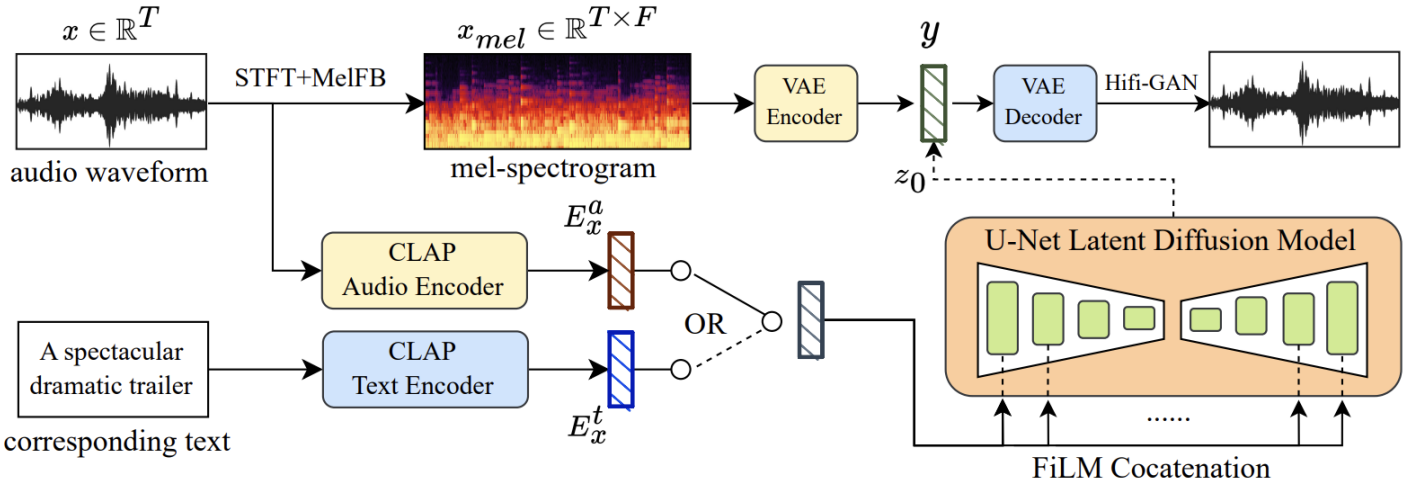


Google MusicFX

https://aitestkitchen.withgoogle.com/ko/tools/music-fx

MusicLDM

- Text-to-music generation model using a Latent Diffusion Model (LDM)
  - Use VAE on spectrogram to obtain latent space and HiFi-GAN for vocoder
  - CLAP (audio-text joint embedding model) for text conditioning



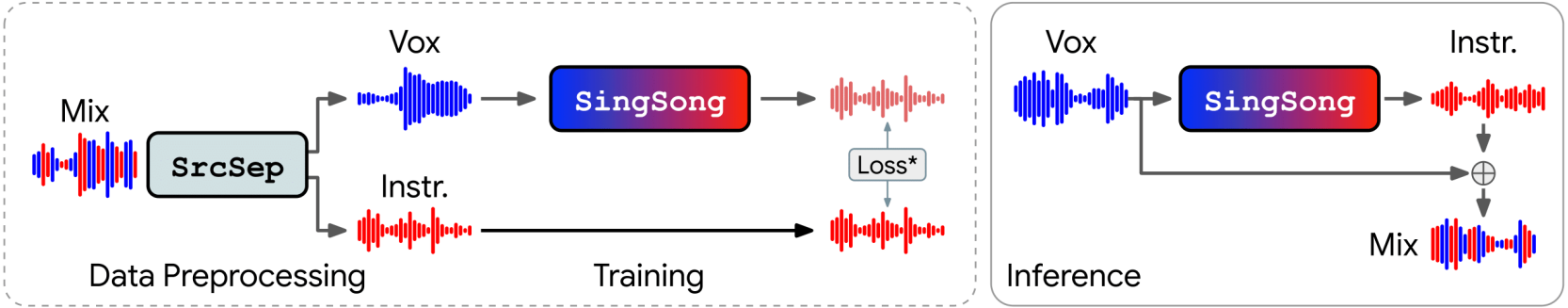
MusicLDM: Enhancing Novelty in Text-to-Music Generation Using Beat-Synchronous Mixup Strategies, Ke Chen, Yusong Wu, Haohe Liu, Marianna Nezhurina, Taylor Berg-Kirkpatrick, Shlomo Dubnov, 2024

MusicLDM

https://musicldm.github.io/

Arrangement Generation

- Generate instrumental sounds given singing voice

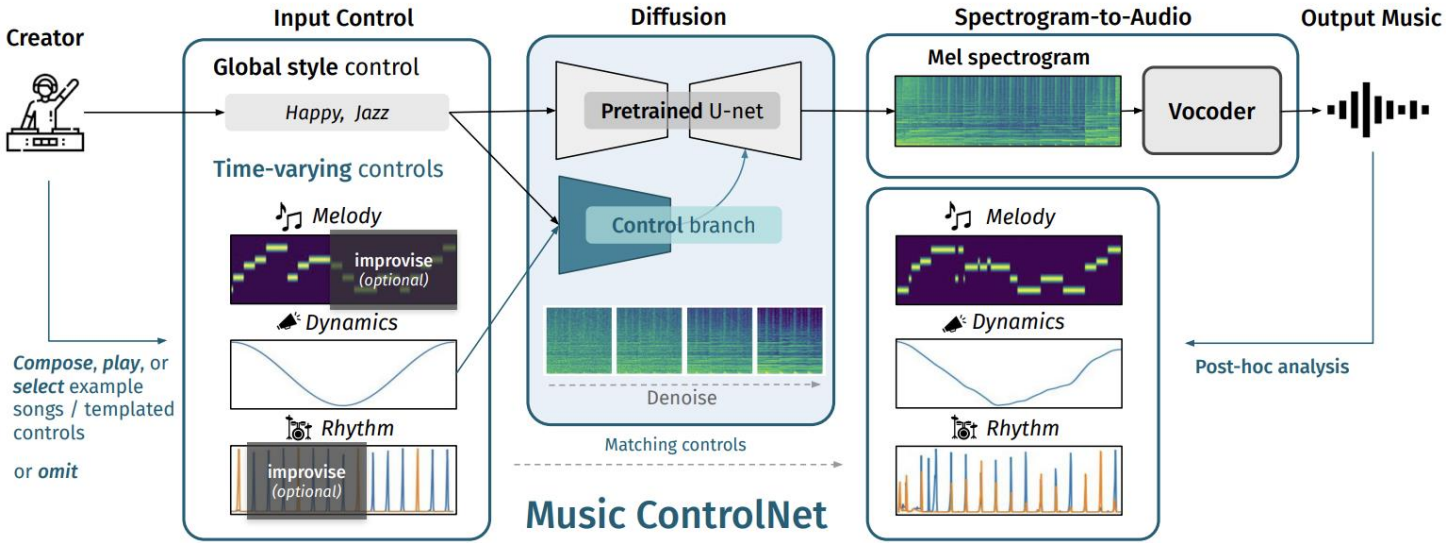


https://storage.googleapis.com/sing-song/index.html

SingSong: Generating musical accompaniments from singing, Chris Donahue, Antoine Caillon, Adam Roberts, Ethan Manilow,…, Arxiv, 2023

Music ControlNet

- Allow temporal music features as a time-varying control
  - Melody (chroma), dynamic curve, beat position



https://musiccontrolnet.github.io/web/

Music ControlNet: Multiple Time-varying Controls for Music Generation, Wu et al, IEEE TASLP, 2024

Music ControlNet

https://www.youtube.com/watch?v=QVr-S-DyccU

Music Generation Services

- Suno: https://suno.com/
- Stable Audio: https://www.stableaudio.com/
- Mix Audio: https://mix.audio/

Concept Video

Music and Audio Computing Lab, KAIST (2018)