PAT 498/598 (Fall 2024)

Special Topics: Generative AI for Music and Audio Creation

Lecture 2: Intro to AI Music

Instructor: Hao-Wen Dong

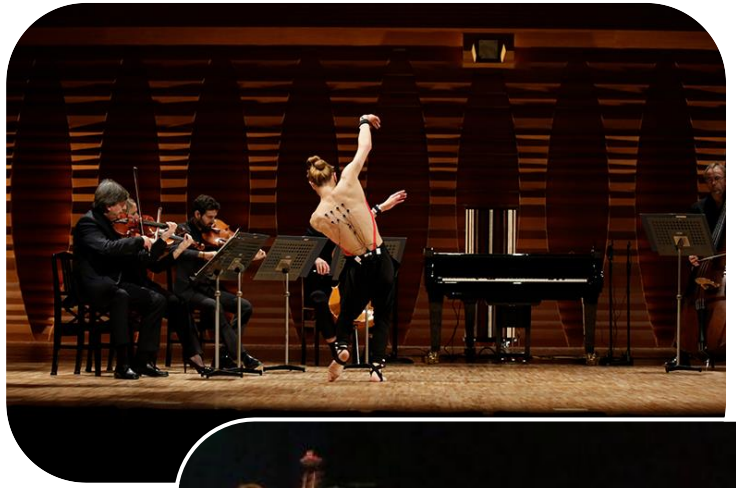
SCHOOL OF MUSIC, THEATRE & DANCE
PERFORMING ARTS TECHNOLOGY
UNIVERSITY OF MICHIGAN

Course Website

- **Main website**: hermandong.com/teaching/pat498_598_fall2024
  - Syllabus, schedule, lecture slides, code examples, etc.
- **Piazza**: Announcements, Q&A
- **Gradescope**: Assignment submission, grading, regrade requests
- **Canvas**: Recordings

(Recap) Music & AI

(Source: Yamaha)



(Source: Robot Gizmos)

(Source: Sankei Shimbun)

(Shlizerman et al., 2019)

(Source: NBC DFW)

Shlizerman et al., "Audio to Body Dynamics," *Proc. CVPR*, 2018.
https://www.yamaha.com/en/news_release/2018/18013101/
https://www.sankei.com/article/20240113-CQCOSQHJWFIYPJJKZDCITRTRVI/
https://www.roboticgizmos.com/shimon-musical-robot-deep-learning/
https://www.nbcdfw.com/entertainment/the-scene/how-verdigris-ensemble-is-using-ai-to-create-a-new-concert-experience/3366031/

(Recap) State of the Art – Text-to-Music Synthesis

Prompt: relaxing and smooth jazz played in a stylish cafe

Prompt: delightful country music with acoustic guitars

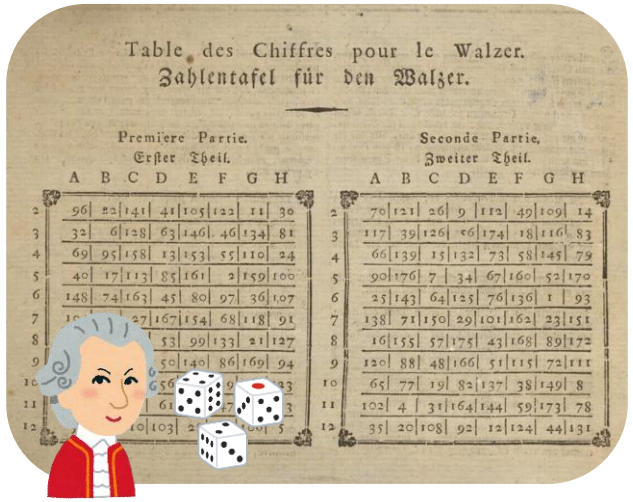
Prompt: cinematic and suspenseful orchestral music

huggingface.co/spaces/facebook/MusicGen

Made with Facebook MusicGen (huggingface.co/spaces/facebook/MusicGen)

(Recap) The Early Days

Musical Dice Game (1792)



(Source: gbrachetta)

gbrachetta.github.io/Musical-Dice/

ILLIAC Suite (1957)

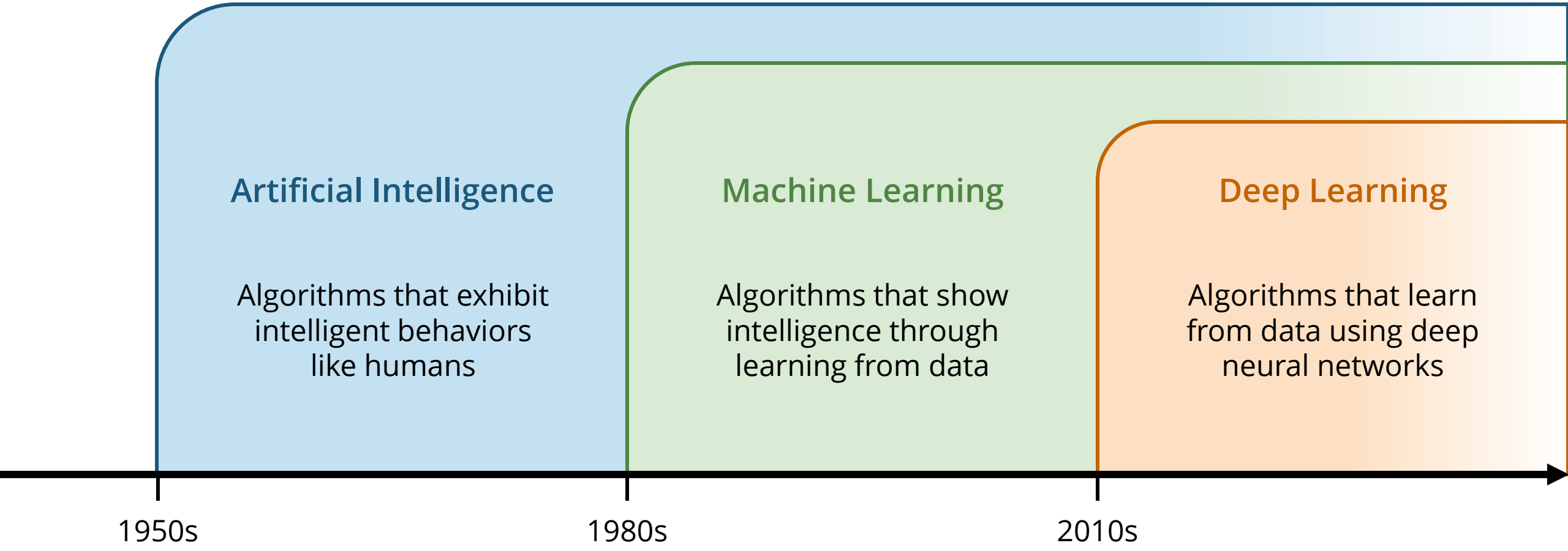
(Source: Illinois Distributed Museum)

Emily Howell (2003)

(Source: The Guardian)

https://gbrachetta.github.io/Musical-Dice/
https://distributedmuseum.illinois.edu/exhibit/illiac-suite/
https://www.theguardian.com/technology/2010/jul/11/david-cope-computer-composer

(Recap) AI vs ML vs DL



Tentative Schedule

Generative AI Background

Week	Date	Lecture
1	Aug 26	Introduction
		Background
	Aug 28	├ AI & machine learning fundamentals
2	~~Sep 2~~	├ ~~No Class (Labor Day)~~
	Sep 4	├ Deep learning fundamentals I
3	Sep 9	├ Deep learning fundamentals II
	Sep 11	├ Language models – RNNs, LSTMs & transformers
4	Sep 16	├ Generative adversarial nets & diffusion models
	Sep 18	└ Music & audio processing fundamentals

Symbolic Music Generation

Week	Date	Lecture
		Symbolic Music Generation
5	Sep 23	├ Melody generation
	Sep 25	├ Harmony & chord progression generation
6	Sep 30	├ Polyphonic music generation
	Oct 2	├ Multitrack music generation
7	Oct 7	├ Multimodal music generation I
	Oct 9	└ Multimodal music generation II
8	~~Oct 14~~	~~No Class (Fall Study Break)~~

Audio Synthesis

Week	Date	Lecture
		Audio Synthesis
	Oct 16	├ Time-domain audio synthesis I
9	Oct 21	├ Time-domain audio synthesis II
	Oct 23	├ Frequency-domain audio synthesis I
10	Oct 28	├ Frequency-domain audio synthesis II
	Oct 30	├ Multimodal audio synthesis I
11	Nov 4	└ Multimodal audio synthesis II
	Nov 6	Project pitch & discussion
12	~~Nov 11~~	~~No Class (Travel)~~
	~~Nov 13~~	~~No Class (Travel)~~

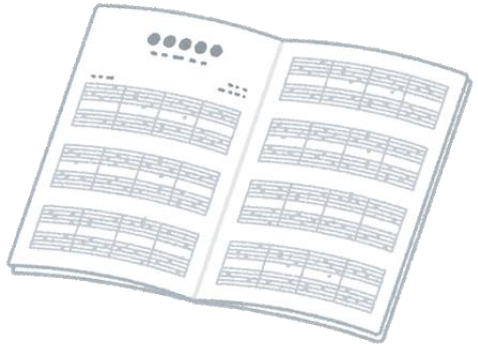
Assistive Music Creation Tools

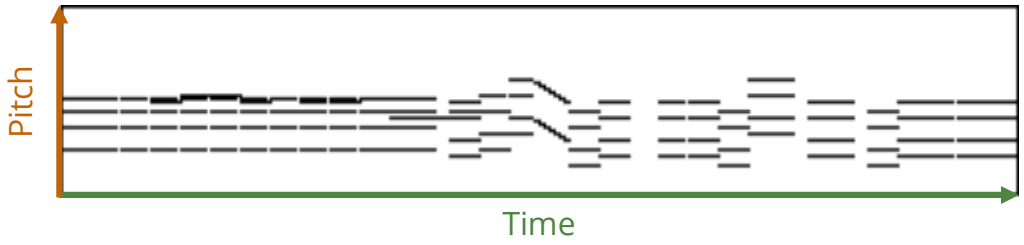
Week	Date	Lecture
		Assistive Music Creation Tools
13	Nov 18	├ Neural audio effects
	Nov 20	├ Auto-mixing
14	Nov 25	└ Live performance & interactive systems
	~~Nov 27~~	~~No Class (Thanksgiving)~~
15	Dec 2	Discussions — ethical concerns & copyright issues
	Dec 4	Review
16	Dec 9	**Project presentation**

Symbolic Music Generation

What is Considered Symbolic Music?

- Sheet music, music notations, pianorolls
- MIDI, MIDI-like notations, text-based notations





Pianoroll

```
Note_on_67, Time_shift_quarter_note, Note_off_67,
Note_on_67, Time_shift_quarter_note, Note_off_67,
Note_on_64, Time_shift_quarter_note, Note_off_64,
Note_on_64, Time_shift_quarter_note, Note_off_64, ...
```

MIDI-like notation

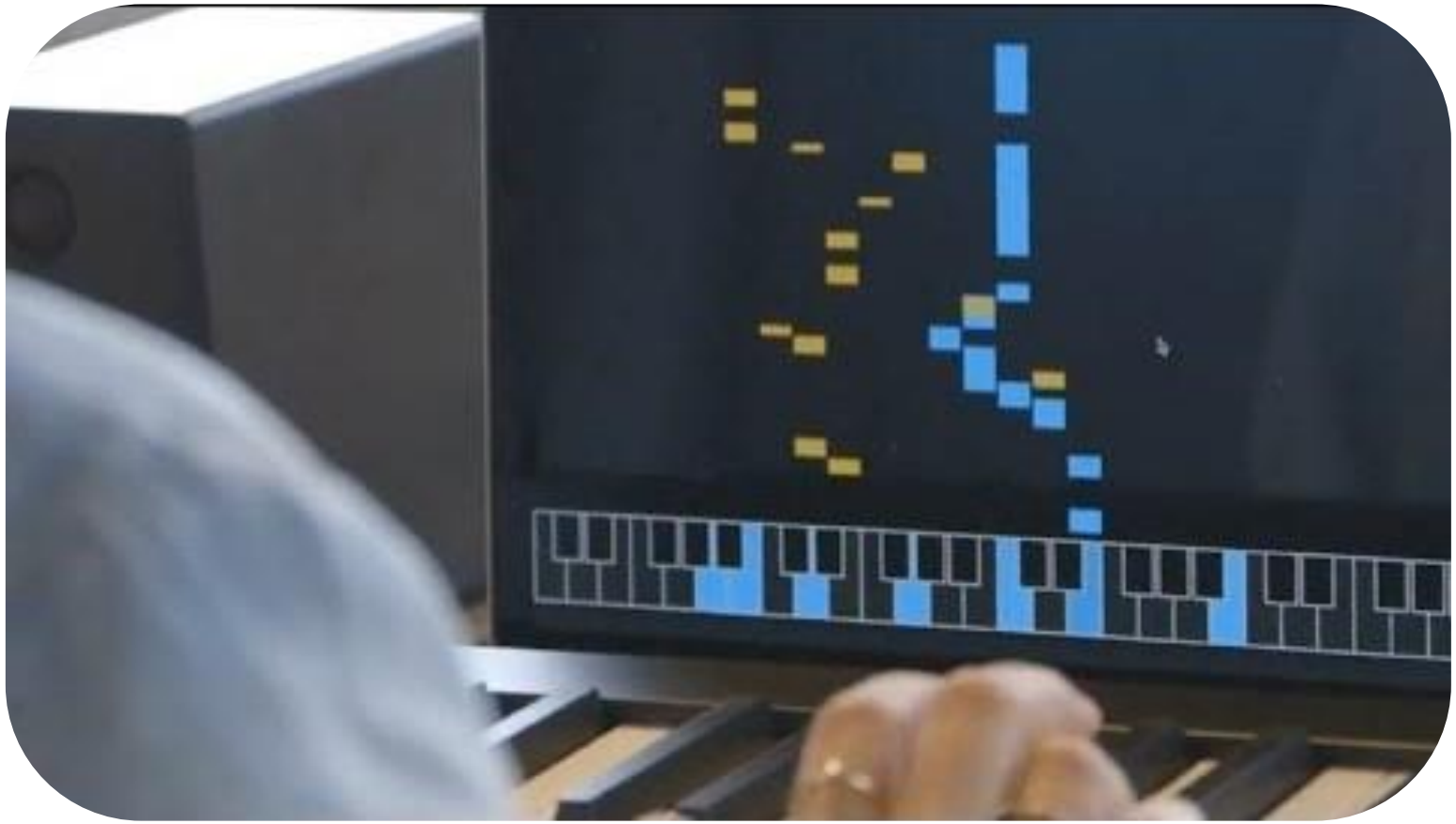
```
X:571
T:Ah! vous dirai-je, maman
T:(Twinkle, twinkle, little star)
C:anon.
O:France
R:Nursery song
M:C
L:1/4
Q:120
K:C
CCGG|AAG2|FFEE|DDC2:|
|:GGFF|EED2|GGFF|EED2|
CCGG|AAG2|FFEE|DDC2:|
```

ABC notation

Topics We’ll Cover

- Melody generation
- Harmony & chord generation
- Polyphonic music generation
- Multimodal music generation

A.I. Duet (2016)



youtu.be/0ZE1bfPtvZo

Piano Genie (2018)

piano-genie.glitch.me/

youtu.be/YRb0XAnUpIk & magenta.tensorflow.org/pianogenie

Chris Donahue, Ian Simon, Sander Dieleman, "Piano Genie," IUI, 2019.

Piano Genie (2018)

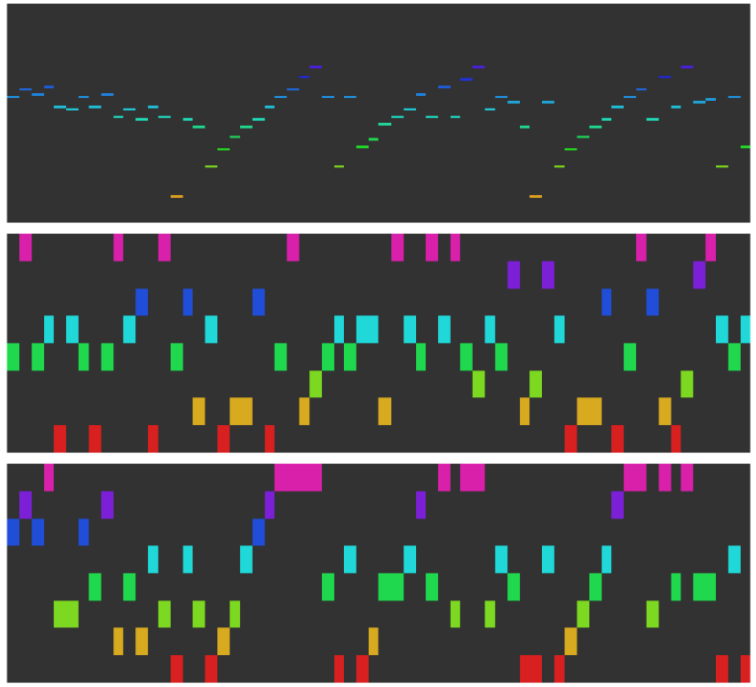
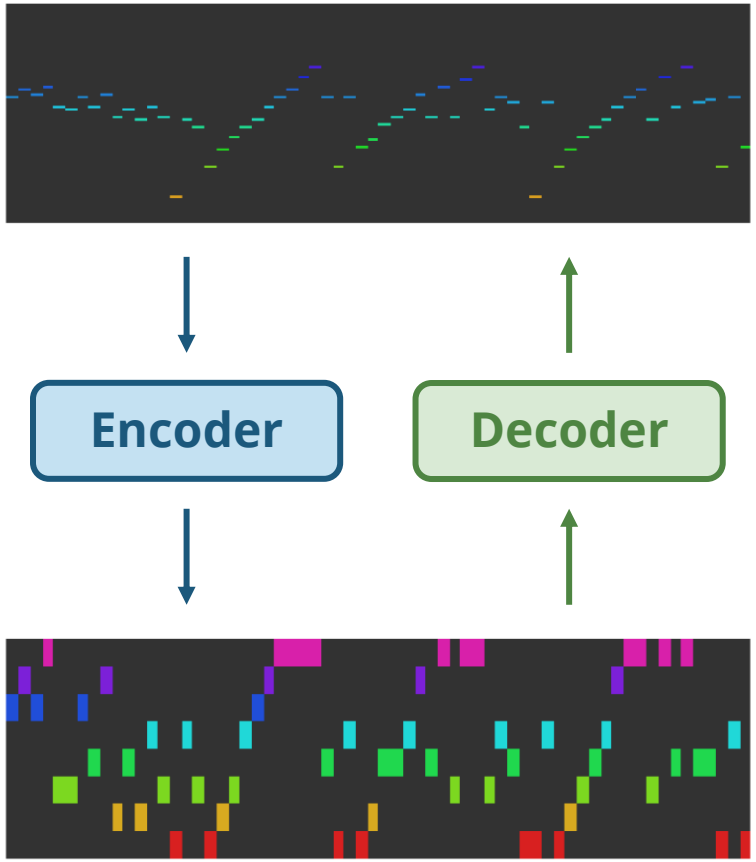


Figure 4: Qualitative comparison of the 8-button encodings for a given melody (top) by the VQ-VAE (middle) and our IQAE with L_{contour} (bottom). Horizontal is note index. The encoding learned by the IQAE echoes the contour of the musical input.

(Donahue et al., 2019)



Chris Donahue, Ian Simon, and Sander Dieleman, "Piano Genie," IUI, 2019.

Fruit Genie (2019)

youtu.be/HoVs4kC68no & magenta.tensorflow.org/fruitgenie

Fruit Genie Live (2019)

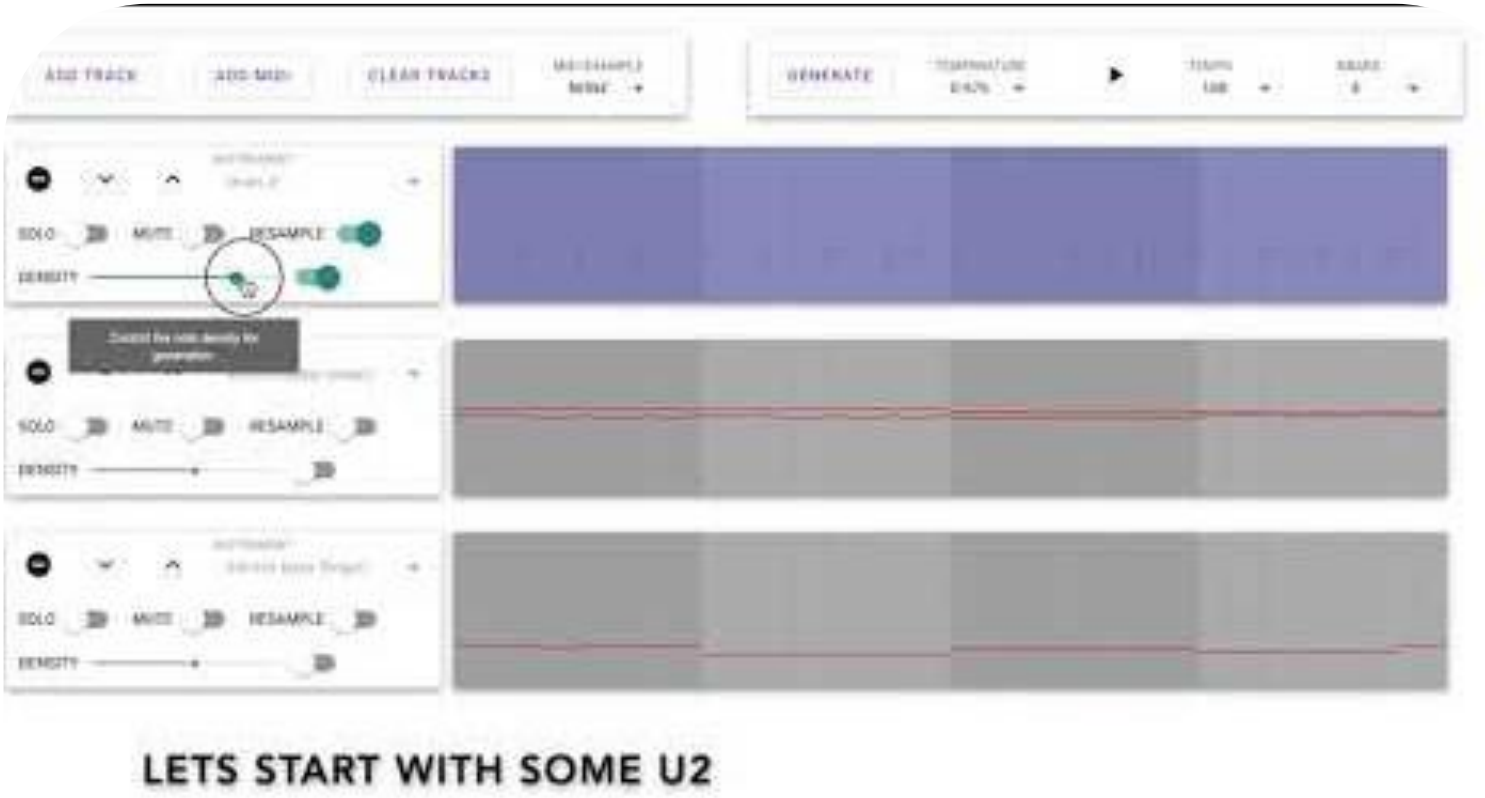
youtu.be/L4wvXrPmIkU & magenta.tensorflow.org/fruitgenie

JS Bach Doodle (2019)

doodles.google/doodle/celebrating-johann-sebastian-bach/

youtu.be/XBfYPp6KF2g & magenta.tensorflow.org/coconet

Multitrack Music Machine (MMM) (2020)

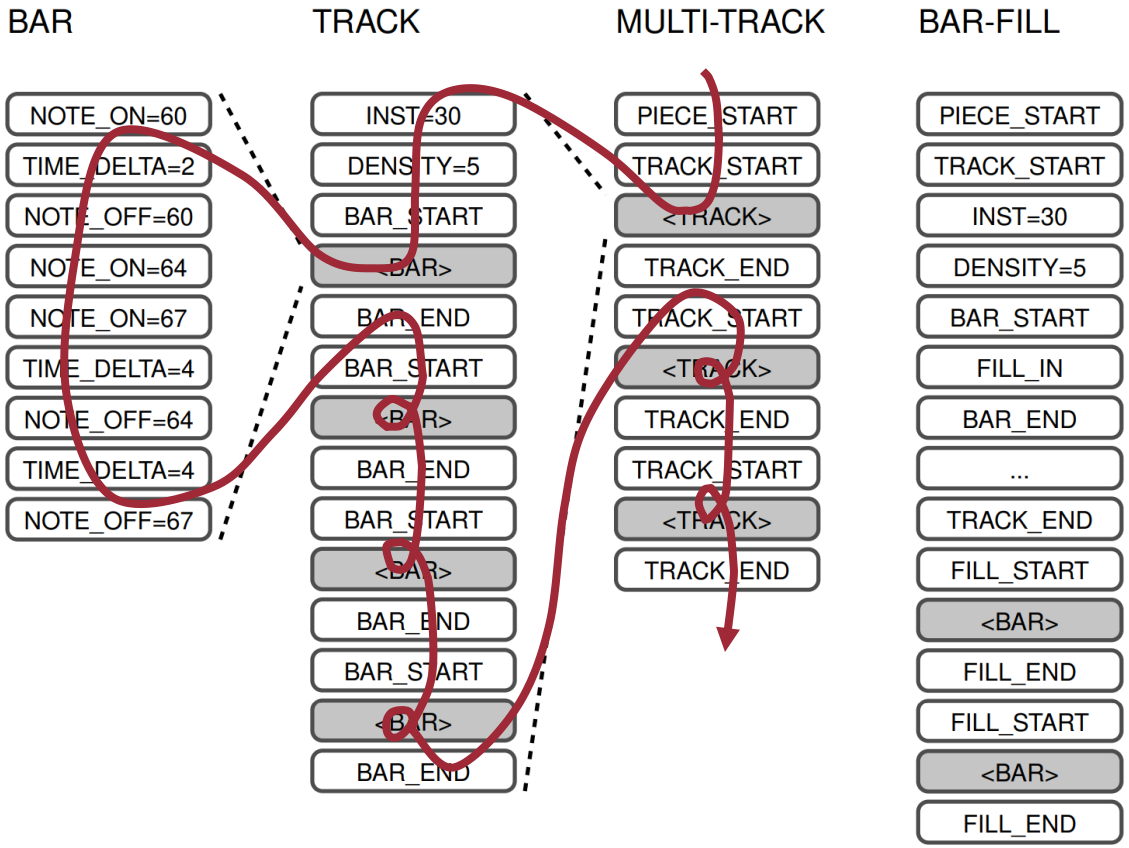


colab.research.google.com/drive/1gzM5Fw2pyWqTqlnkOKlUWO91WCg1m3LJ

youtu.be/NdeMZ3y-84Q &
metacreation.net/projects/mmm-multi-track-music-machine

Jeff Ens and Philippe Pasquier, "MMM : Exploring Conditional Multi-Track Music Generation with the Transformer," *arXiv preprint arXiv:2008:06048*, 2020.

Multitrack Music Machine (MMM) (2020)



(Ens & Pasquier, 2020)

Fig. 1. The MultiTrack and BarFill representations are shown. The `<bar>` tokens correspond to complete bars, and the `<track>` tokens correspond to complete tracks.

Jeff Ens and Philippe Pasquier, "MMM : Exploring Conditional Multi-Track Music Generation with the Transformer," *arXiv preprint arXiv:2008:06048*, 2020.

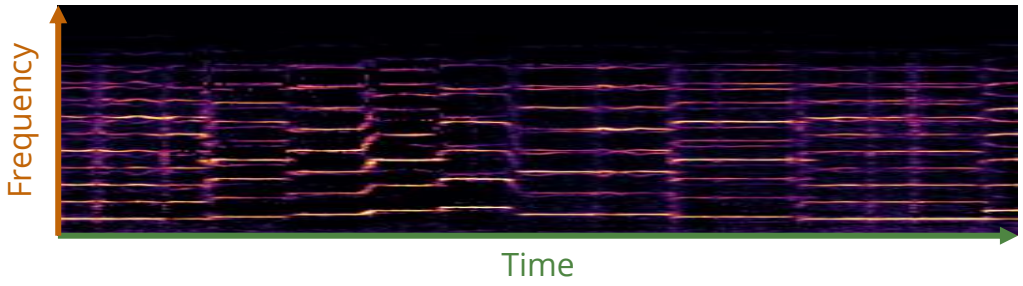
MMM4Live (2024)

youtu.be/KtxEESFjy2g & metacreation.net/projects/mmm4live

Audio Synthesis

Topics We'll Cover

- Time-domain audio synthesis
- Frequency-domain audio synthesis
- Multimodal audio synthesis

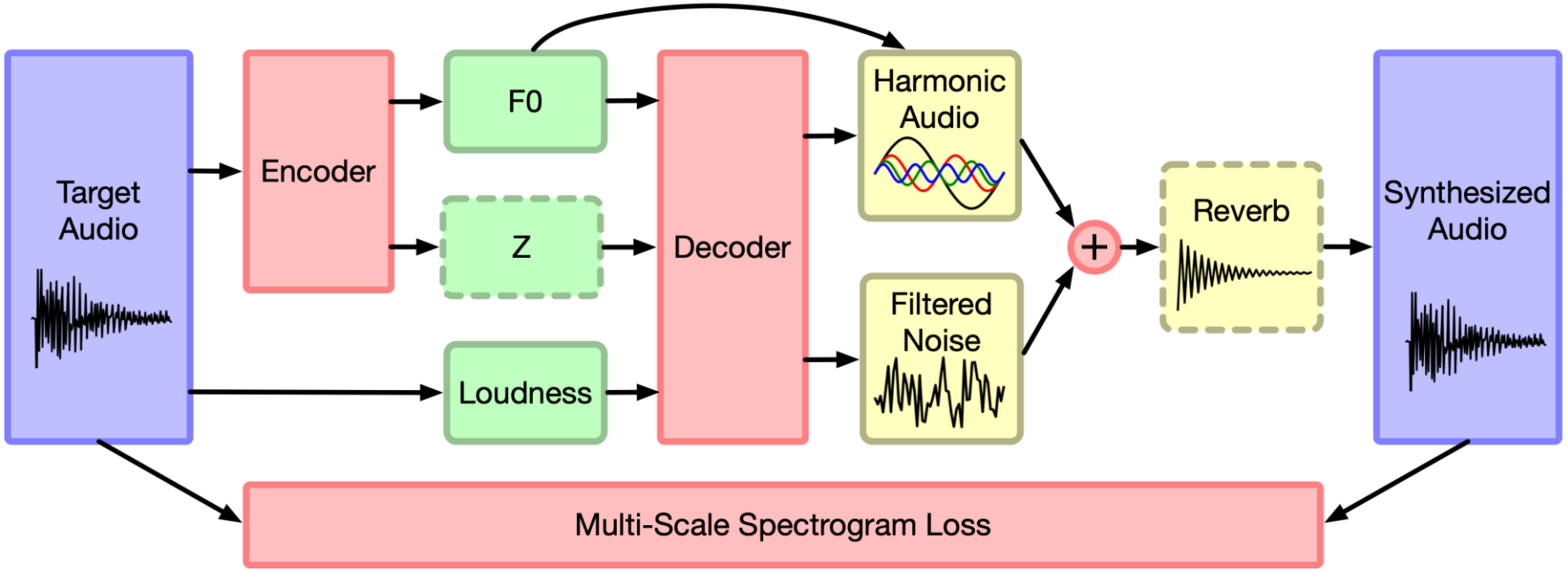


Tone Transfer (2020)

sites.research.google/tonetransfer

youtu.be/bXBliLjImio & magenta.tensorflow.org/tone-transfer

Differentiable Digital Signal Processing (DDSP) (2020)

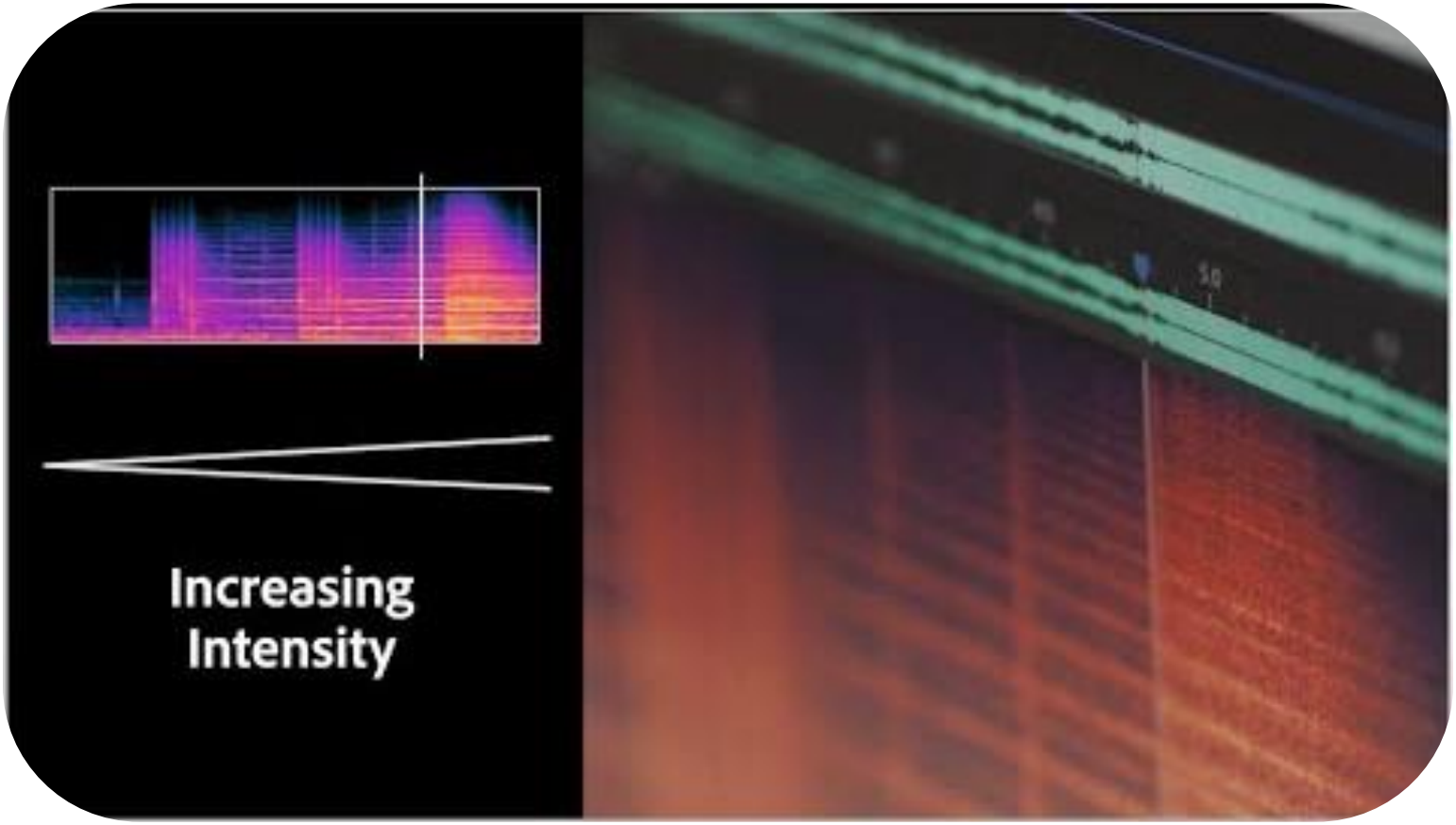


Jesse Engel, Lamtharn (Hanoi) Hantrakul, Chenjie Gu, Adam Roberts, "DDSP: Differentiable Digital Signal Processing," *ICLR*, 2020.

Adobe Firefly for Videos (2023)

youtu.be/30xueN12guw

Music GenAI Project at Adobe (2024)



youtu.be/YRb0XAnUpIk

DITTO: Diffusion for Music Generation (2024)

youtu.be/KooosSNPNo8 & ditto-music.github.io/web/

DITTO: Diffusion for Music Generation (2024)

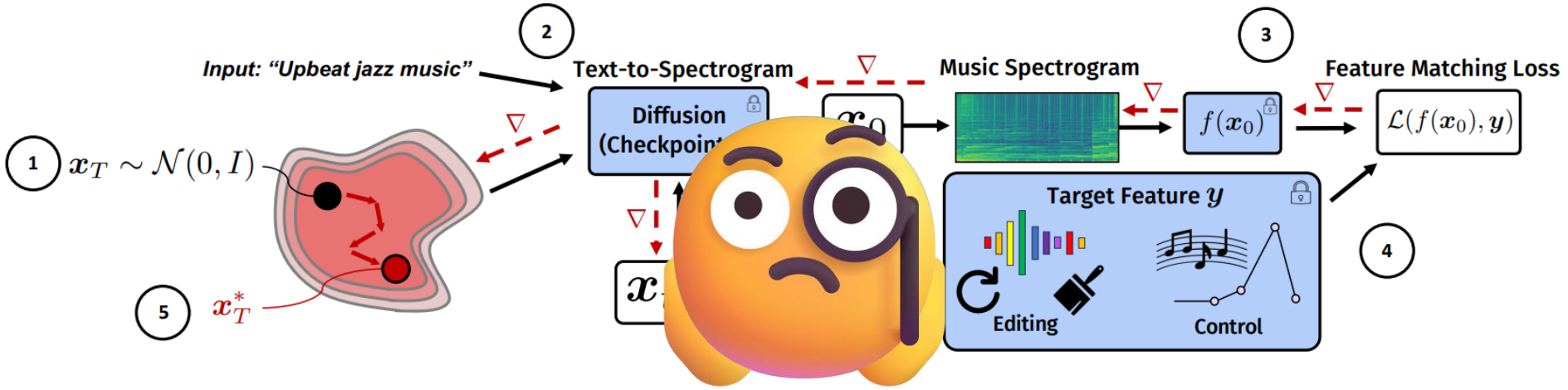


Figure 1. We propose **DITTO**, or **D**iffusion **I**nference-**T**ime $\boldsymbol{T}$-**O**ptimization, a general-purpose framework to control pre-trained diffusion models at inference-time. 1) We sample an initial noise latent $\boldsymbol{x}_T$; 2) run diffusion sampling to generate a music spectrogram $\boldsymbol{x}_0$; 3) extract features from the generated content; 4) input a target control signal; and 5) optimize the initial noise latent to fit any differentiable loss.

At the end of this course, you should be comfortable reading such a figure.

Zachary Novack, Julian McAuley, Taylor Berg-Kirkpatrick, and Nicholas J. Bryan, "DITTO: Diffusion Inference-Time T-Optimization for Music Generation," *ICML*, 2024.

Assistive Music Creation Tools

Assistive Music Creation Tools

- Any tools used in the music creation pipeline
- Some tasks are well-developed (e.g., auto-tune)
- Counterintuitively, **assistive tools are often time more difficult to build**
  - For it requires **FIRST understanding the input** and **THEN perform the task**

Topics We’ll Cover

- Neural audio effects
- Auto-mixing
- Live performance & interactive systems

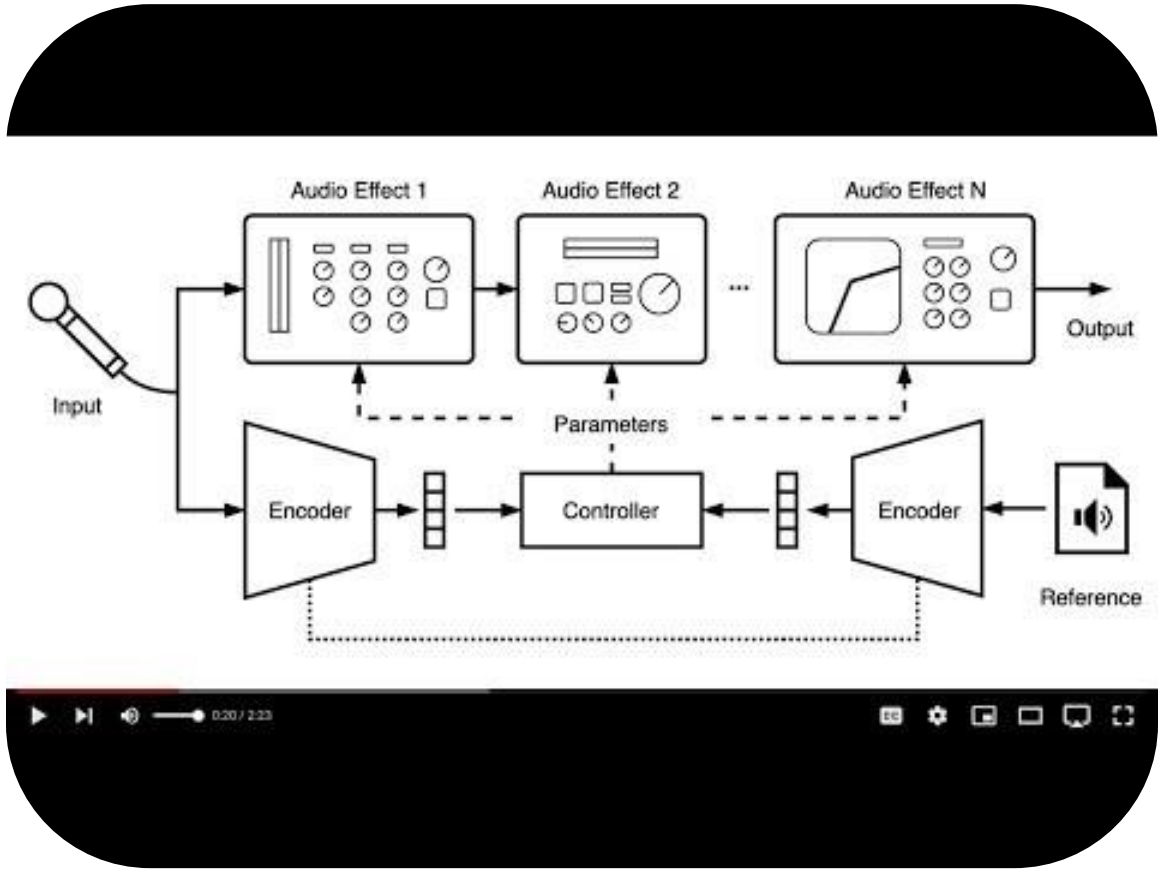
Neural Audio Effects (2021)

youtu.be/Zmo8kB-SfF4

colab.research.google.com/github/csteinmetz1/steerable-nafx/blob/master/steerable-nafx.ipynb

Christian J. Steinmetz and Joshua D. Reiss, "Steerable discovery of neural audio effects," *NeurIPS ML4CD Workshop*, 2021.

DeepAFx-ST: Style Transfer of Audio Effects (2022)



youtu.be/lZp455wiMk4?t=100

Christian J. Steinmetz, Nicholas J. Bryan, and Joshua D. Reiss, "Style Transfer of Audio Effects with Differentiable Signal Processing," *JAES*, 2022.

DeepAFx-ST: Style Transfer of Audio Effects (2022)

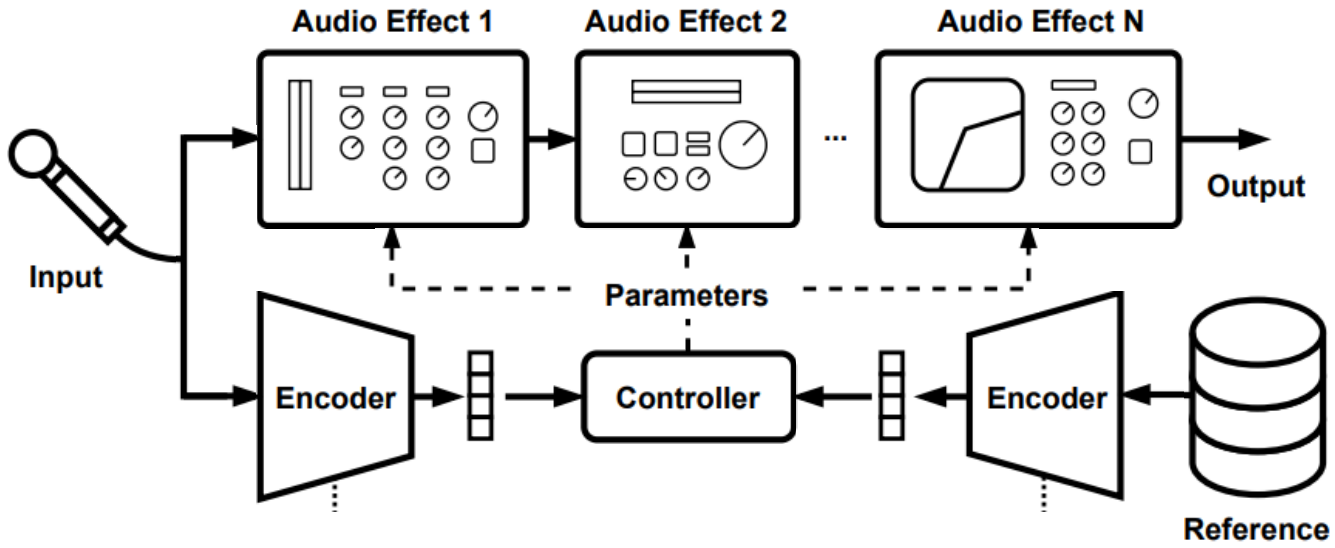


Fig. 1. Our *DeepAFx-ST* method imposes the audio effects and production style from one recording to another by example. We use a shared-weight encoder to analyze the input and a style reference signal, then compare each with a controller that outputs the parameters of effects that themselves perform style manipulation.

Christian J. Steinmetz, Nicholas J. Bryan, and Joshua D. Reiss, "Style Transfer of Audio Effects with Differentiable Signal Processing," *JAES*, 2022.

Realtime Neural Audio Synthesis – RAVE (2022)

youtu.be/dMZs04TzxUI

github.com/acids-ircam/RAVE

Antoine Caillon and Philippe Esling, "RAVE: A variational autoencoder for fast and high-quality neural audio synthesis," *arXiv preprint arXiv:2111.05011*, 2021.

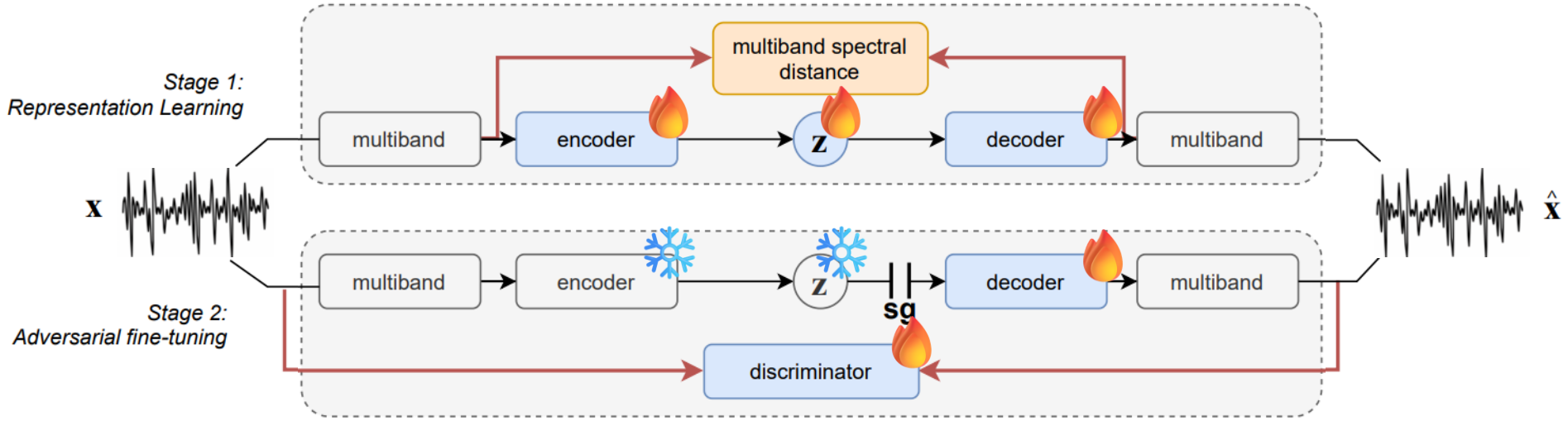
Realtime Neural Audio Synthesis – RAVE (2022)

youtu.be/jAlRf4nGgYI

github.com/acids-ircam/RAVE

Antoine Caillon and Philippe Esling, "RAVE: A variational autoencoder for fast and high-quality neural audio synthesis," *arXiv preprint arXiv:2111.05011*, 2021.

Realtime Neural Audio Synthesis – RAVE (2022)



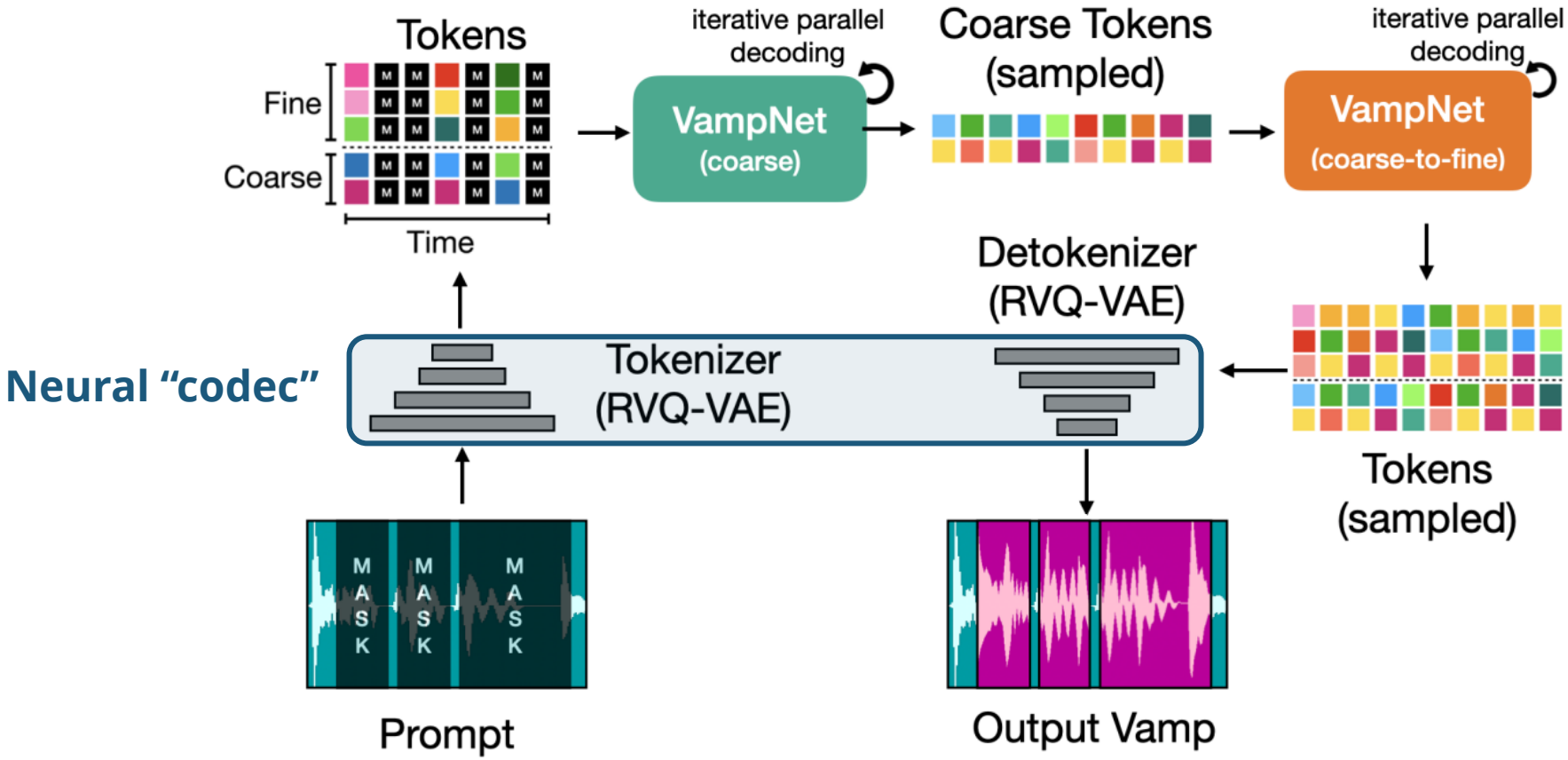
Antoine Caillon and Philippe Esling, "RAVE: A variational autoencoder for fast and high-quality neural audio synthesis," *arXiv preprint arXiv:2111.05011*, 2021.

unloop: a looper that doesn't repeat itself (2023)

youtu.be/yzBI8Vcjd2s

github.com/hugofloresgarcia/unloop

VampNet



Hugo Flores Garcia, Prem Seetharaman, Rithesh Kumar, and Bryan Pardo, "VampNet: Music Generation via Masked Acoustic Token Modeling," *ISMIR*, 2023.

Human-AI Music Co-creativity

AI Song Contest

- Annual international competition showcasing the **creative potential of human–AI co-creativity in the songwriting process**

aisongcontest.com

Yaboi Hanoi – Entering Demons & Gods (2022)

https://youtu.be/PbrRoR3nEVw

soundcloud.com/yaboihanoi/enter-demons-and-gods

Reading: The Making of Entering Demons & Gods (2022)

“**It was like a saxophonist trained in classical Thai motifs, who played a special ‘Thai Edition’ saxophone with Phi Nai tunings, had joined the musical conversation.** The same was true with the trumpet model and the ขลุ่ย ‘Khlui’ - a flute from Thai, Laos and Cambodian repertoire. I could assemble a **transcultural ensemble** to expand the sonic palette of Thai motifs, whilst adhering to underlying tunings and idiomatic inflections like never before.”

lamtharnhantrakul.github.io/enter-demons-and-gods/

Synthetic Beat Brigade - How would you touch me? (2023)

youtu.be/O4cJ3acEGDw & bit.ly/45vlmuT

Synthetic Beat Brigade - How would you touch me? (2023)

- **Ideation**: Spotify API, ChatGPT, Facebook Llama, Google Bison
- **Lyrics**: ChatGPT 2, Genius API
- **Composition**: AI Drummachine, Mofi, Tonetrasnfer, This patch does not exist, Albeatz, BaiscPitch, Magenta, AIVA, MuseNet
- **Vocals**: Soundly Voice Designer, Vocal Remove, Voice characteristics
- **Mastering**: Landr
- **Cover art, bandart**: Midjourney
- **Clip**: ComfyUI for Stable Diffusion + ControlNet

Reading: The Making of How would you touch me? (2023)

“This project is a collaboration between **Artificial Intelligence (AI) enthusiasts in four fields: artist management, music and post-production, tech, and creative**. In contrast, the majority of the music industry sees AI as a threat. Our team understands that these technological advances will have a significant impact on how we produce music. Because of this, we have decided to **use AI for every step of the production process**. From ideation to creating the lyrics to producing the music.”

drive.google.com/file/d/1QTQ7P3iZI6I0anIwNQ3ewf8g3JjDjesl/view