

CULTURAL AUTOMATION WITH MACHINE LEARNING

SESSION 05: GENERATIVE AUDIO

Parag K. Mital
UCLA DMA

CULTURAL CONTEXT + HISTORY OF GENERATIVE AUDIO



Still-Life with Chair Caning, 1912 by Pablo Picasso

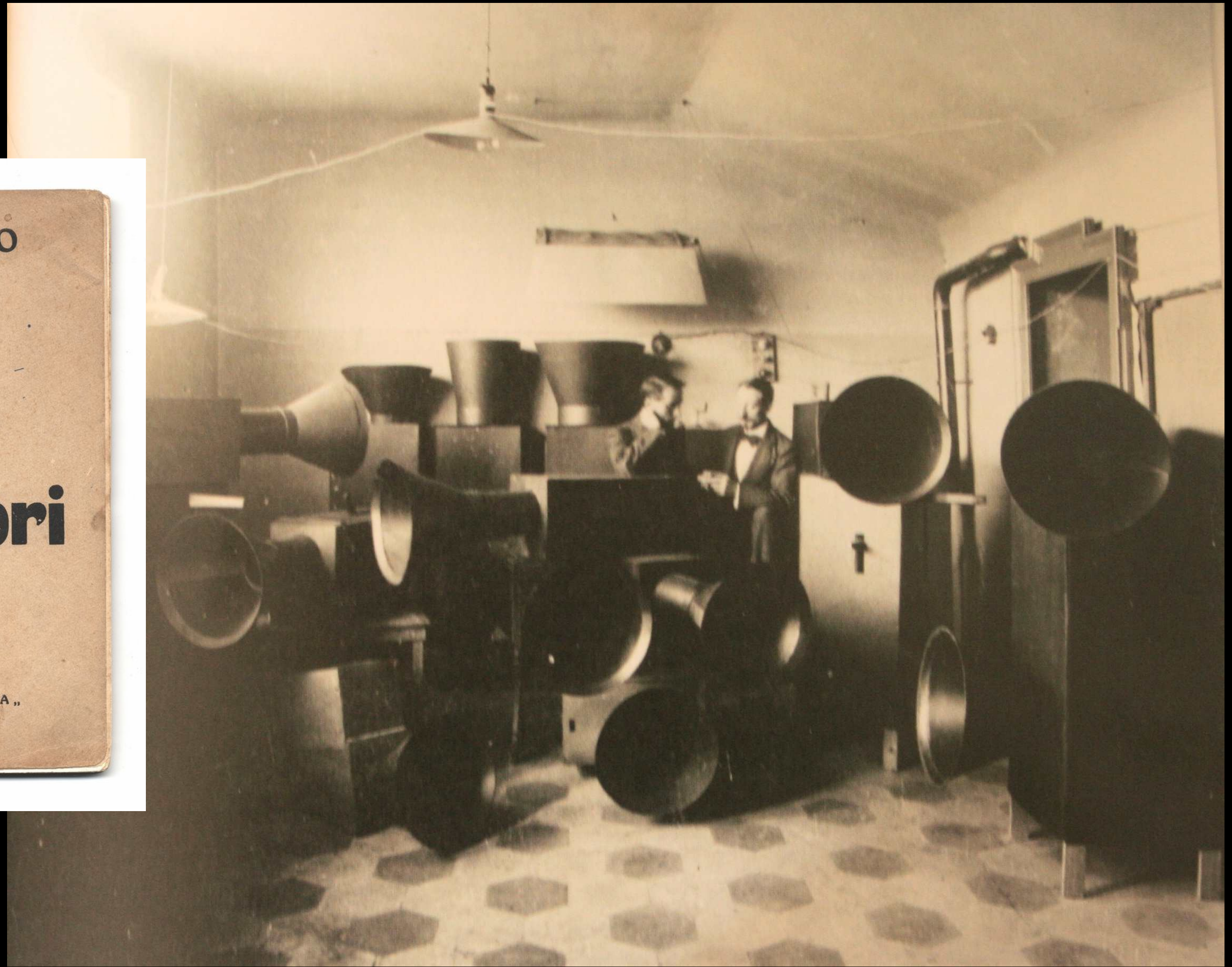


Walter Ruttmann: Weekend (1930)

<https://www.youtube.com/watch?v=v3cEAMh3eng>



Hans Richter



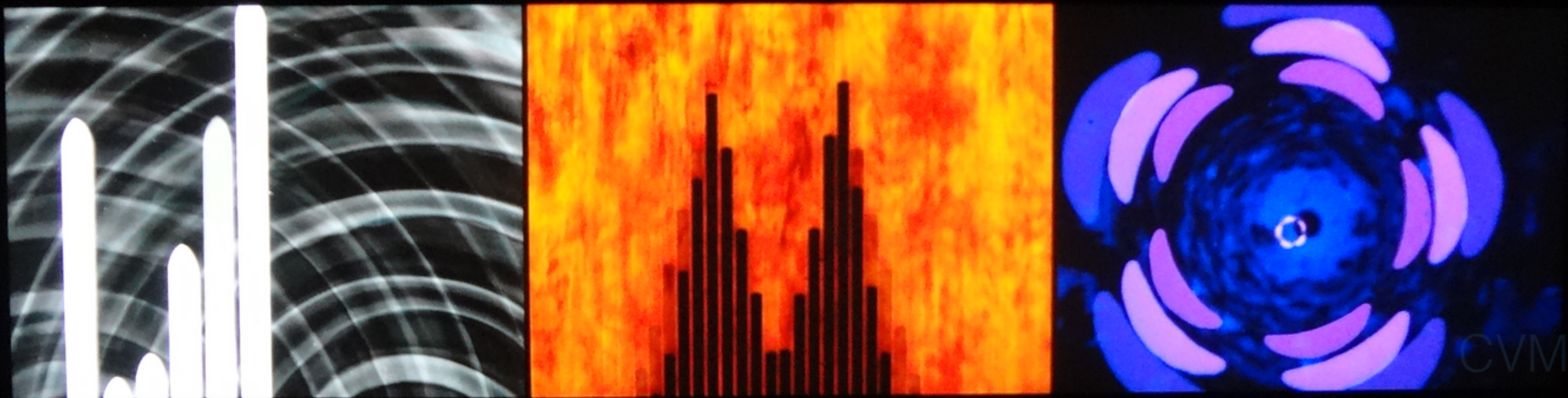
"The Art of Noise"



Pierre Schaeffer - "Musique concrète"



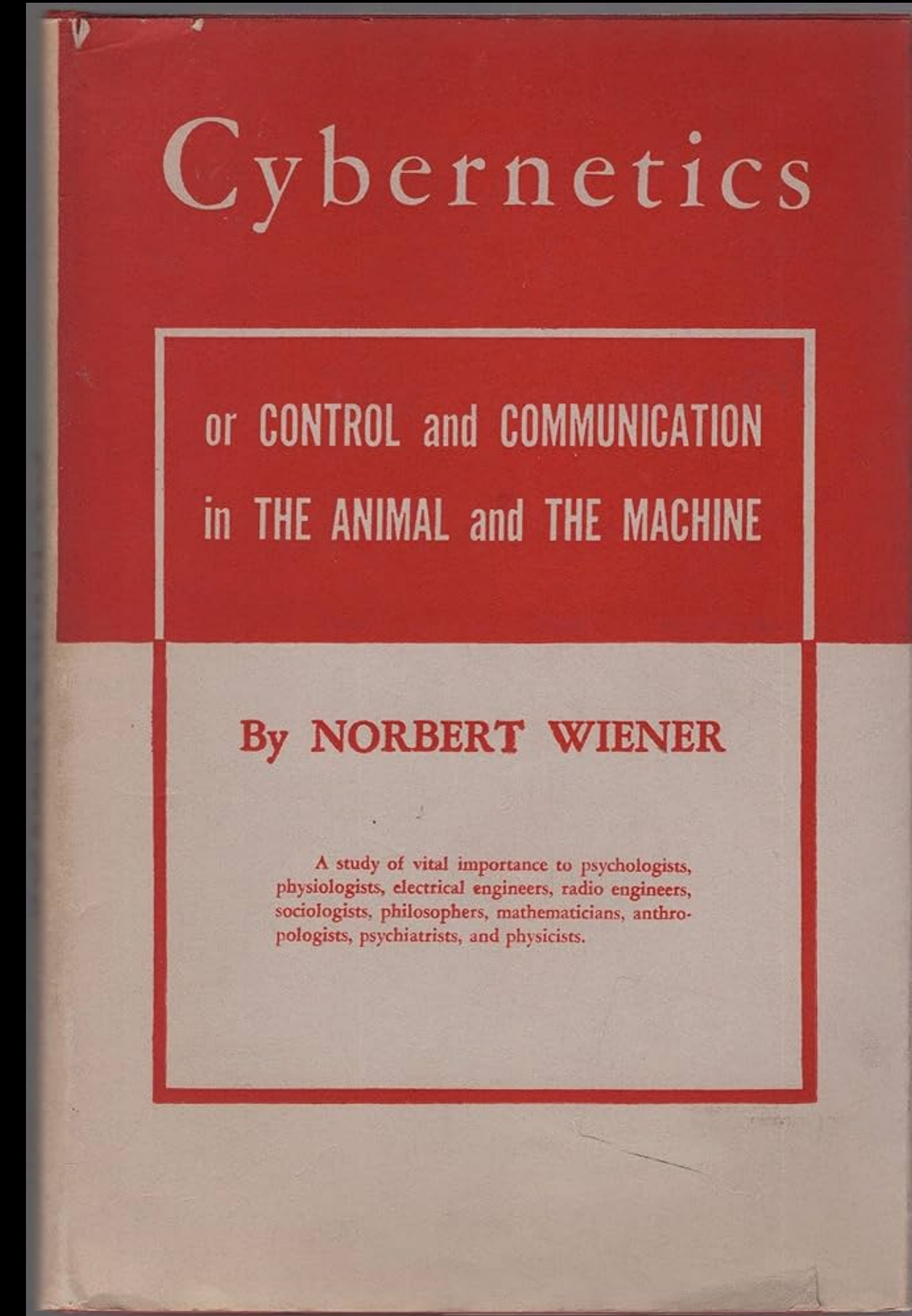
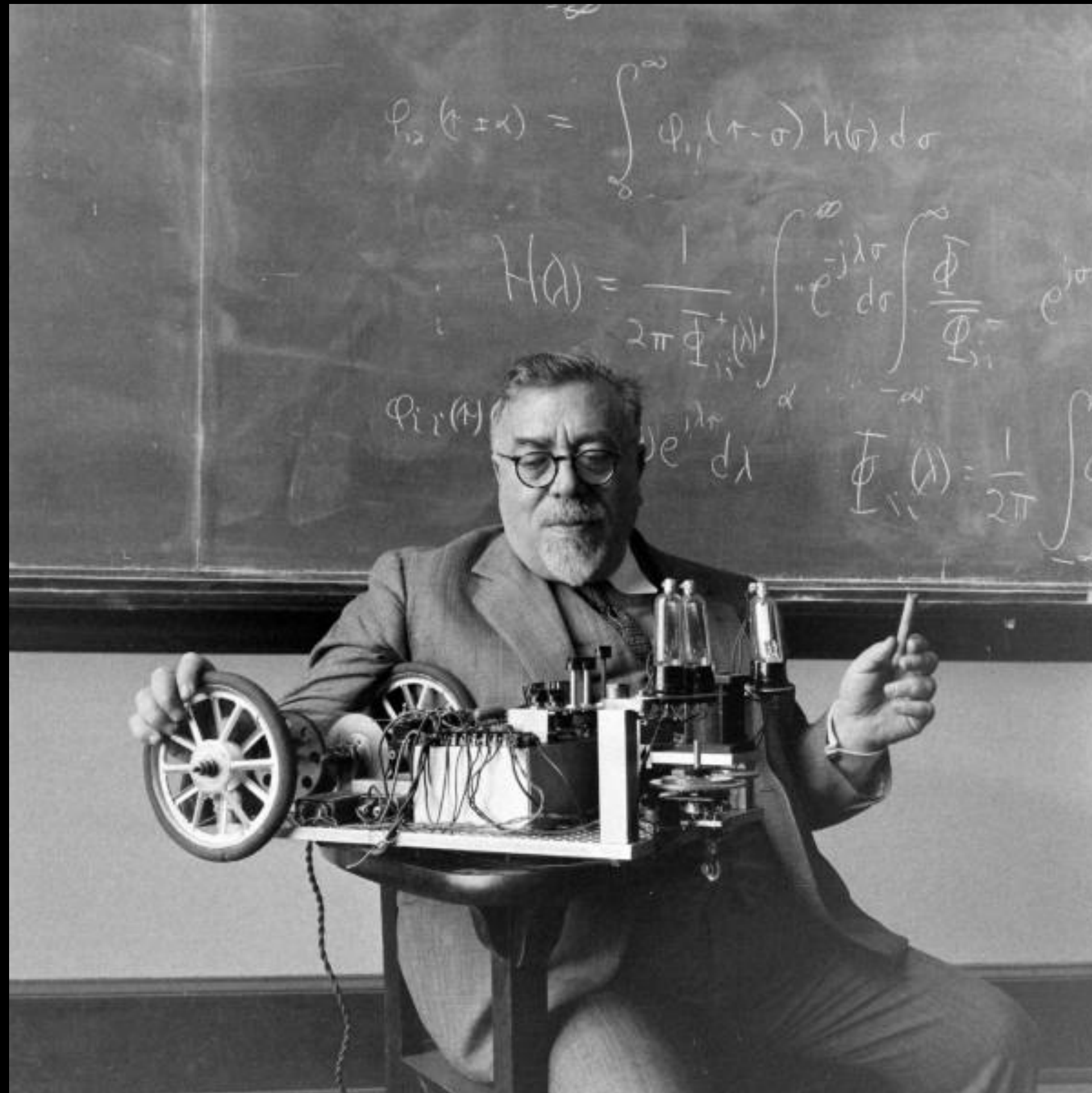
Norman McClaren - visual music



Oskar Fischinger - visual music

<https://www.acmi.net.au/education/school-program-and-resources/the-art-of-visual-music/>

<https://www.google.com/logos/2017/fischinger/fischinger17.html>



Norbert Wiener - cybernetics research and his 1948 book "Cybernetics: Or Control and Communication in the Animal and the Machine"



<https://www.bl.uk/collection-items/earliest-known-recording-of-computer-generated-music#:~:text=In%201951%2C%20a%20BBC%20outside,recording%20of%20computer%2Dgenerated%20music>



https://en.wikipedia.org/wiki/Illiacc_Suite

<https://www.youtube.com/watch?v=fojKZ1ymZlo>



<https://www.musicradar.com/news/tech/a-brief-history-of-computer-music-177299>



Daphne Oram - visual music (1960's)

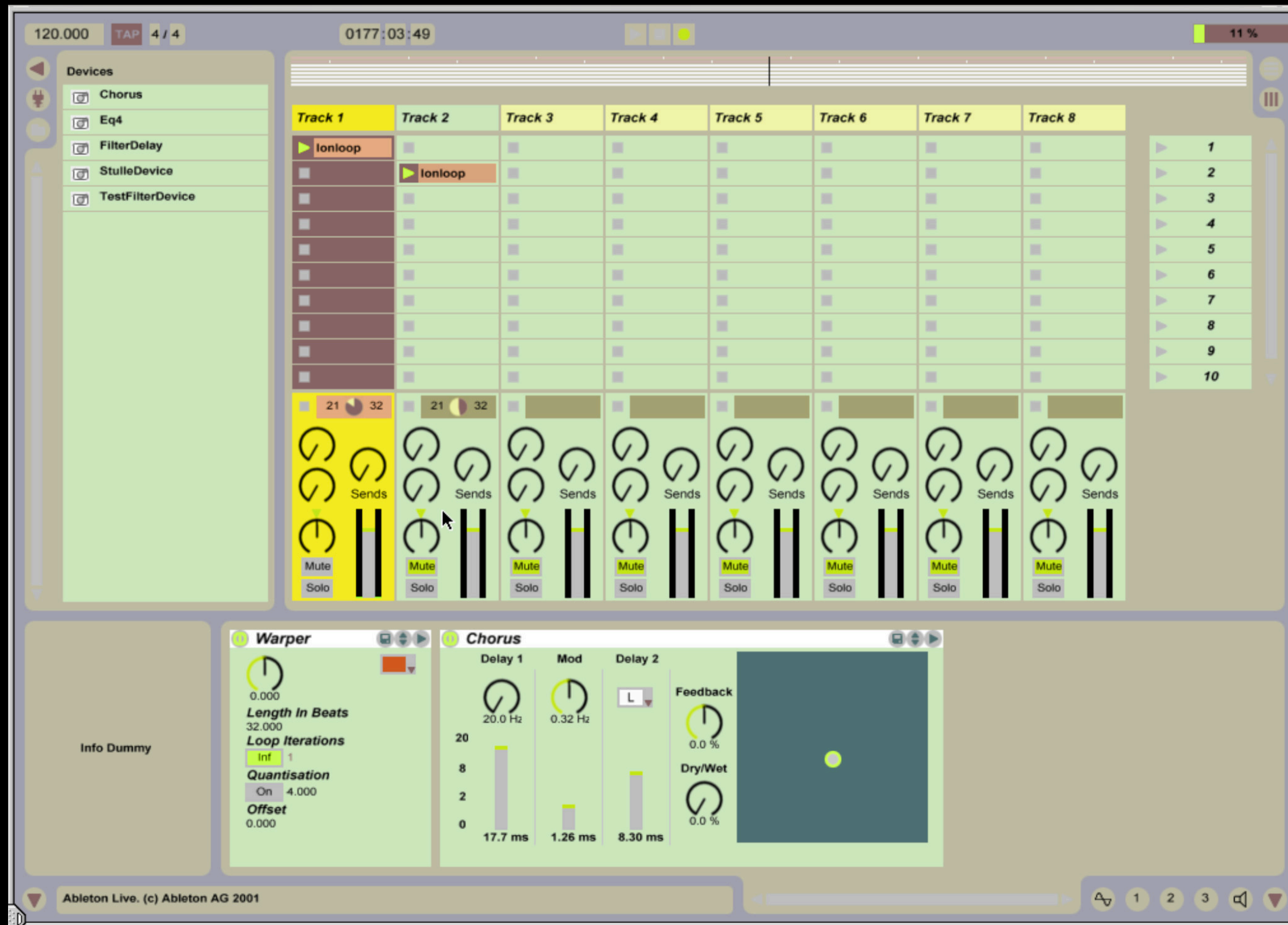
<https://www.daphneoram.org/oramicsmachine/>



David Tudor



Dj Kool Herc



Ableton Live



The Sugarhill Gang - Rapper's Delight (Official Video)

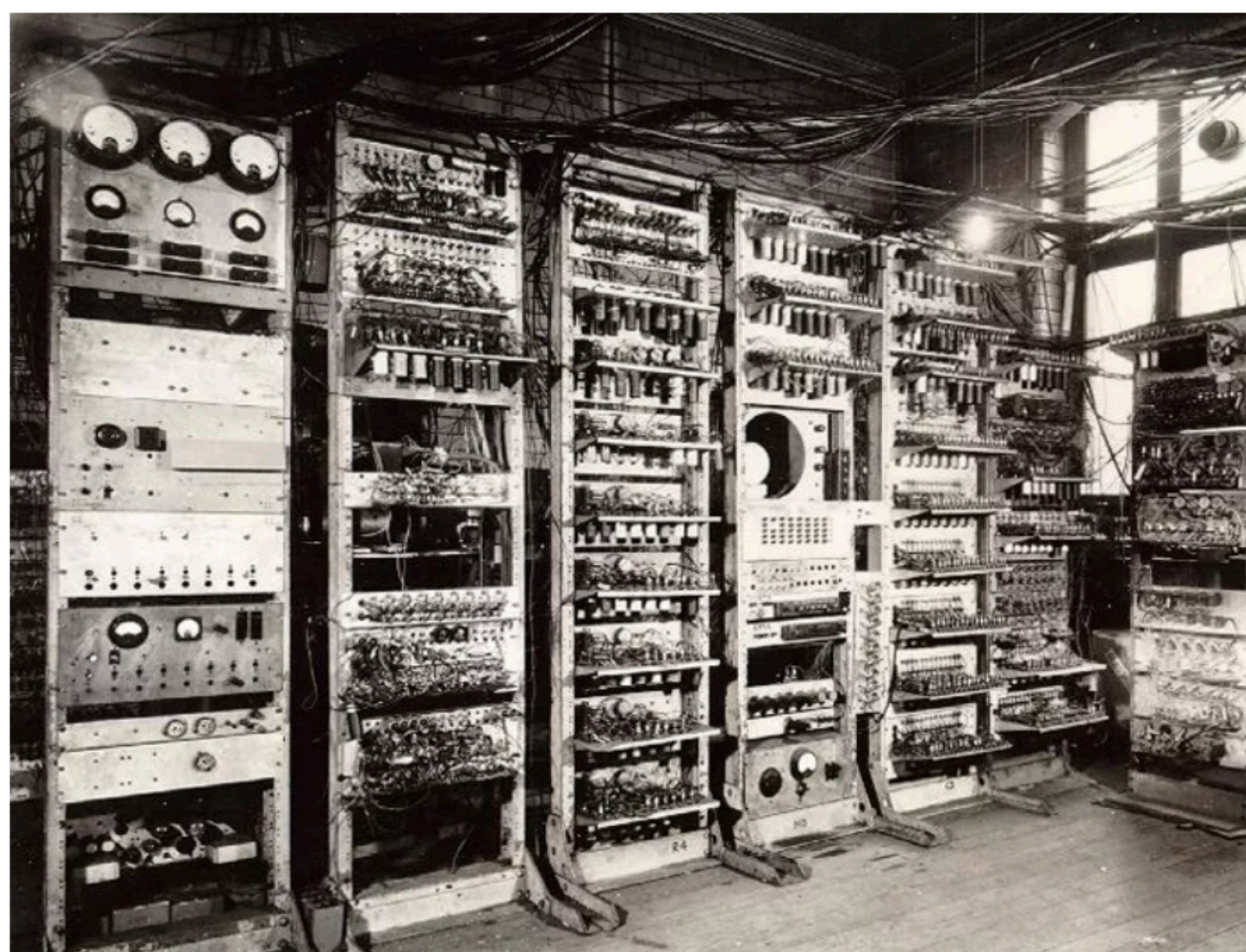
<https://www.youtube.com/watch?v=mcCK99wHrk0>

News

A brief history of computer music

By [Computer Music Specials](#) published October 13, 2008

From Baa Baa Black Sheep to Ableton Live



'The Baby' was the forerunner to the Ferranti Mark 1 that was used to make the first computer music recording.

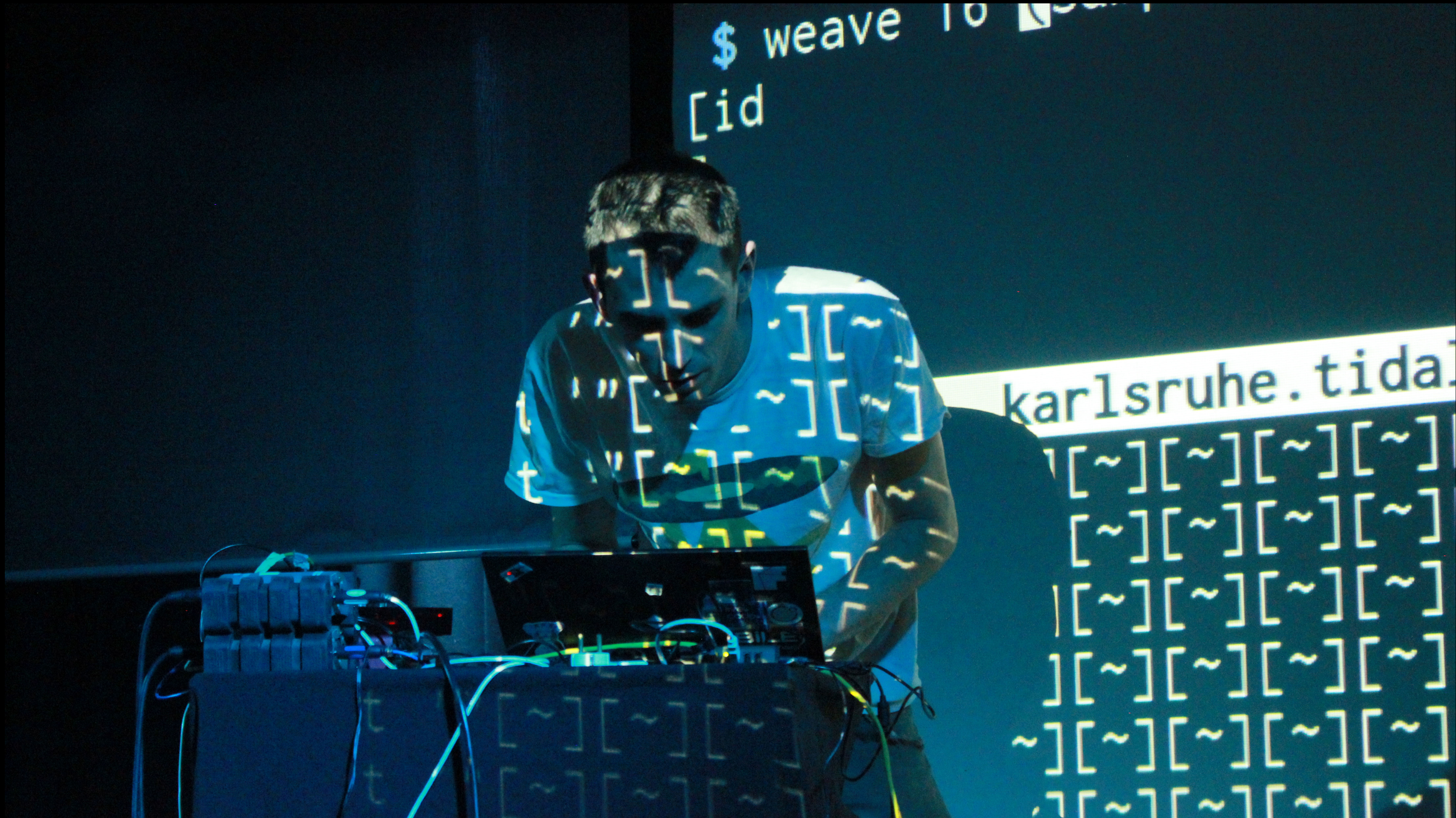
History of Computer Music:

<https://www.musicradar.com/news/tech/a-brief-history-of-computer-music-177299>



rjdj

<https://en.wikipedia.org/wiki/RjDj>





Libraries Resources Downloads Gallery Tutorials Links/Acks

What Is Impromptu?

Impromptu is an OSX programming language and environment for composers, sound artists, VJ's and graphic artists with an interest in live or interactive programming. Impromptu is a Scheme language environment, a member of the Lisp family of languages. Impromptu is used by artist-programmers in livecoding performances around the globe.

Time plays a major role in the Impromptu environment allowing accurate real-time scheduling of events and code. Impromptu is a dynamic environment designed for the creation and manipulation of running programs in live performance.

Impromptu is a programmable AudioUnit host. A powerful environment for creating signal processing graphs of arbitrary complexity with precise programmatic control over individual AU nodes. Musical material can be precisely scheduled for performance by any AudioUnit instrument node and parameters, program changes and presets can be programmatically changed on-the-fly as well as directly via the AU's user interface.

Impromptu allows you to write, modify, compile and hot swap custom signal processing code - an ability that allows you to write low level sound synthesis code during live performance. This functionality is designed to interoperate with the AudioUnit signal graph allowing you to freely mix your favorite AudioUnits with custom synthesis code. All on-the-fly of course.

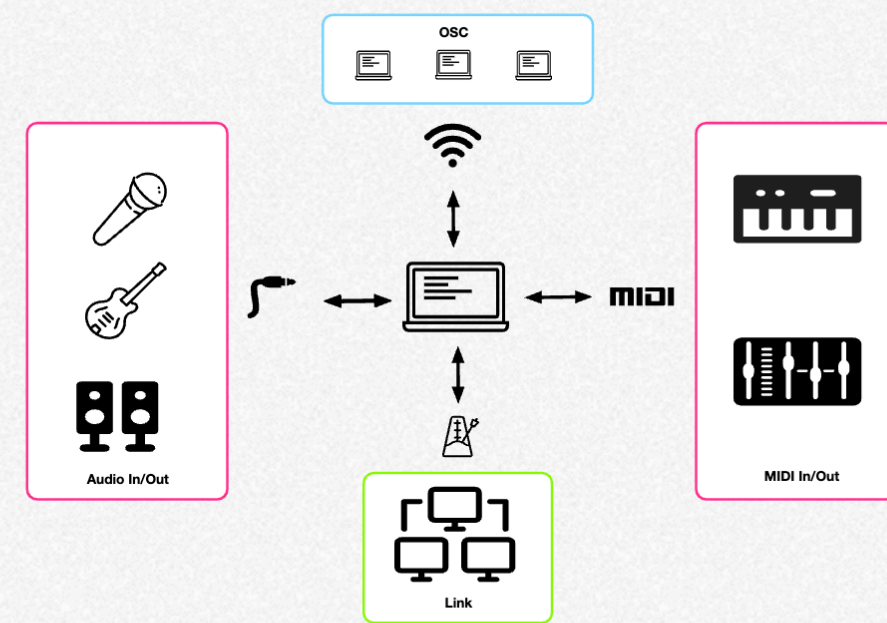
Graphics routines can be applied with the same temporal accuracy as audio material allowing artists to tightly integrate audio and visual components. OpenGL, live video processing, vector drawing routines, image rendering, CoreImage filters, text rendering and quicktime movie support are a few of the graphics features available for artists to play with.

Impromptu also includes a bidirectional ObjC-Bridge allowing Scheme to instantiate and call ObjC objects and ObjC objects to call back into the Scheme interpreter. An LLVM based scheme->x86 compiler is also available supporting on-the-fly compilation of scheme code for audio signal processing, OpenGL, numeric processing and alike.



<http://impromptu.moso.com.au/>

Live Code Everything



Sonic Pi lets you use **simple code** to turn your computer into a fully networked live coding **music studio**:

- Multi Channel **Audio** In/Out
- Well-timed **MIDI** In/Out
- Well-timed **OSC** (Open Sound Control) In/Out
- Ableton's **Link** network metronome built-in



Sonic Pi

*Experience the **sound of code**.*

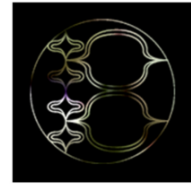
Sonic Pi is your free code-based music creation and performance tool.

- Powerful** for professional musicians and DJs.
- Expressive** for composition and performance.
- Accessible** for blind and partially sighted people.
- Simple** for computing and music lessons.

<https://sonic-pi.net/>

Bronze Player App

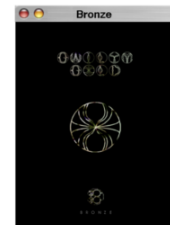
Bronze logo



There are 3 pages for the app: Home, Player and Information. Unless otherwise stated, this shows how the player will work across the iPhone and Desktop versions.

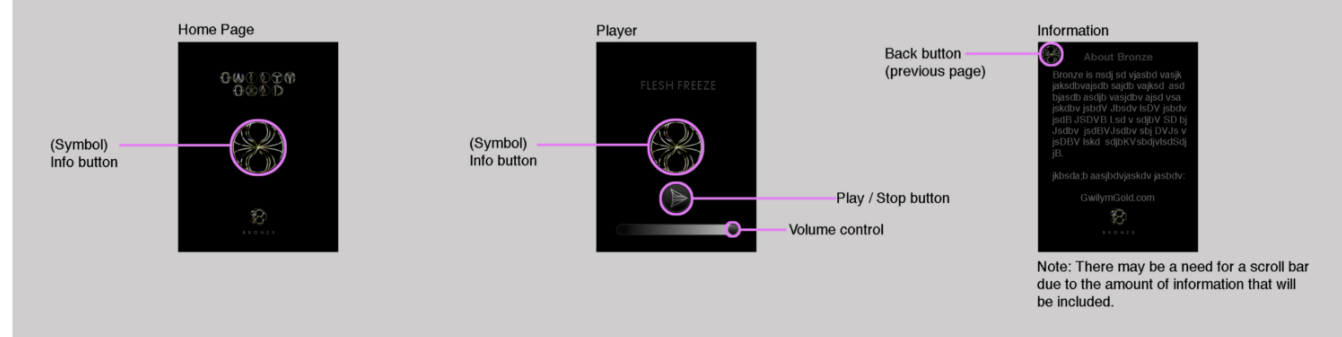
I have included sequences and actions for when the buttons are pressed and the transitions for the pages - if any of these are tricky to implement across the iPhone and / or Desktop versions or if there is an easier way, then it would be best to go down that route.

The information page included only gives a rough outline as to how the page will eventually look.



Note: For the desktop version of the app, a title bar will need to be included

Interactive Elements



1. The home page fades in from black.

Appears for a few seconds, then the Gwynn Gold logo and Bronze Player (app) logo fade to black and the player page fades in.

The symbol, in the middle of the app, remains throughout.

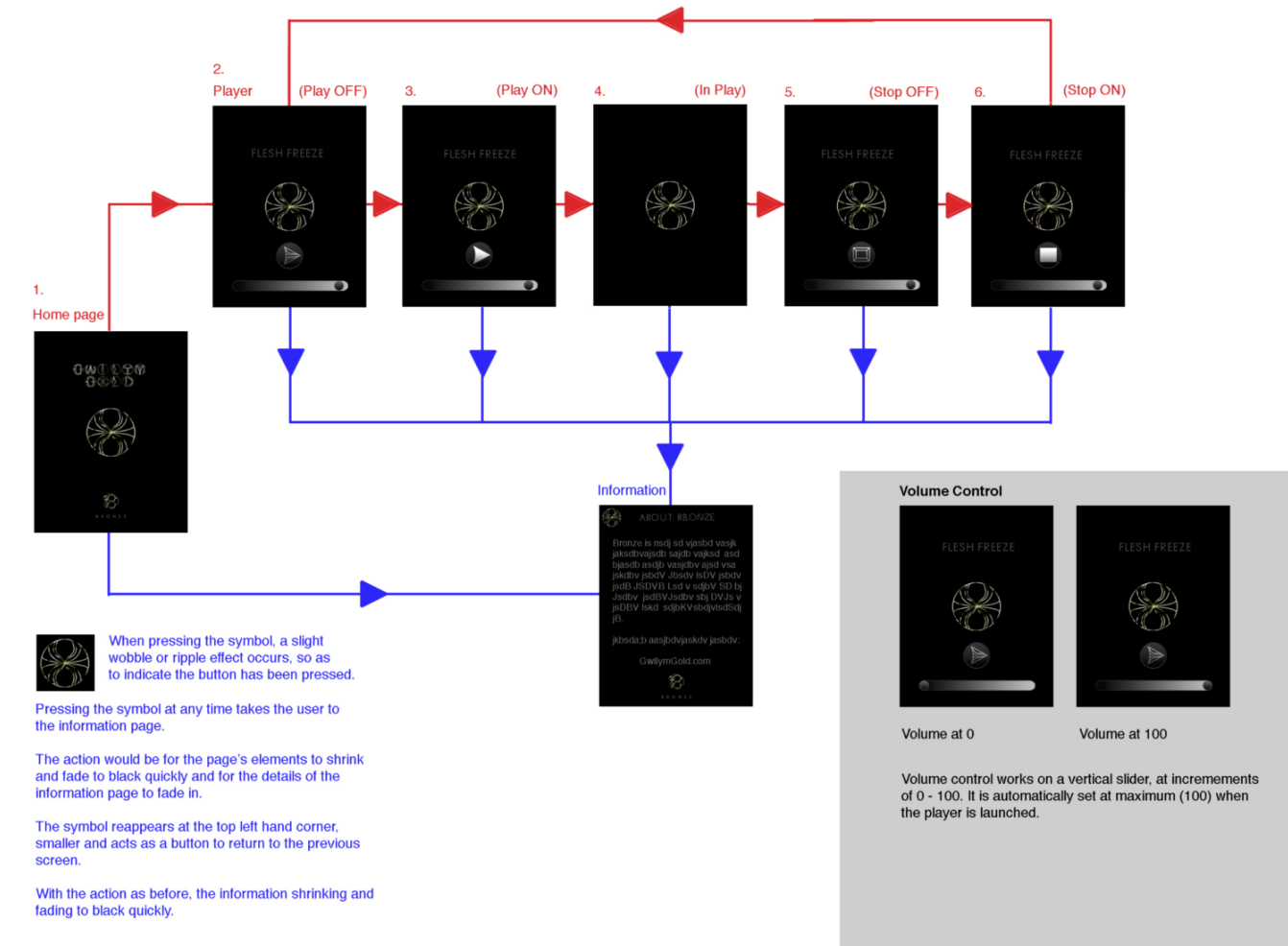
2. The song title, Play / Stop button and volume fade in. The player is set in Play OFF mode. Volume is set at maximum.

3. Pressing the play button, the arrow switches from an outline to a filled symbol. The play button is ON.

4. Once playing, the controls; Play / Stop button, the track title and volume all fade to black. The symbol remains.

5. Touching the screen (Pod app) or moving your cursor (desktop app) brings the controls back, fading in from black. The play button has changed to stop.

6. Pressing the stop button, the outline for the button becomes the filled symbol. Just as before, after a few seconds, the button changes to the play button in OFF mode (2).



When pressing the symbol, a slight wobble or ripple effect occurs, so as to indicate the button has been pressed.

Pressing the symbol at any time takes the user to the information page.

The action would be for the page's elements to shrink and fade to black quickly and for the details of the information page to fade in.

The symbol reappears at the top left hand corner, smaller and acts as a button to return to the previous screen.

With the action as before, the information shrinking and fading to black quickly.



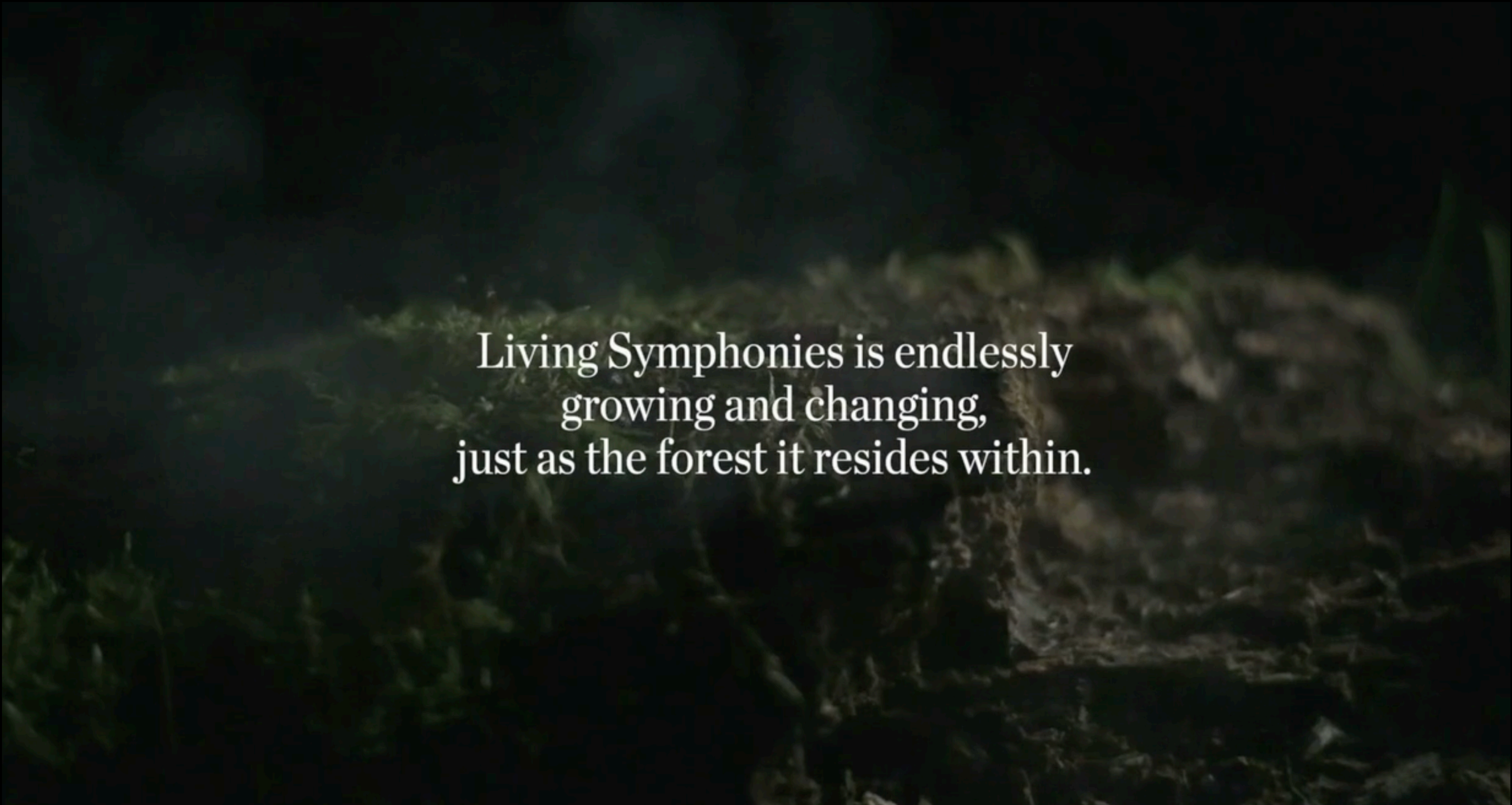
Bronze Format

<https://pkmital.com/home/works/bronze-format/>



Memory Mosaic

<https://pkmital.com/home/works/memory-mosaic/>



Living Symphonies is endlessly
growing and changing,
just as the forest it resides within.

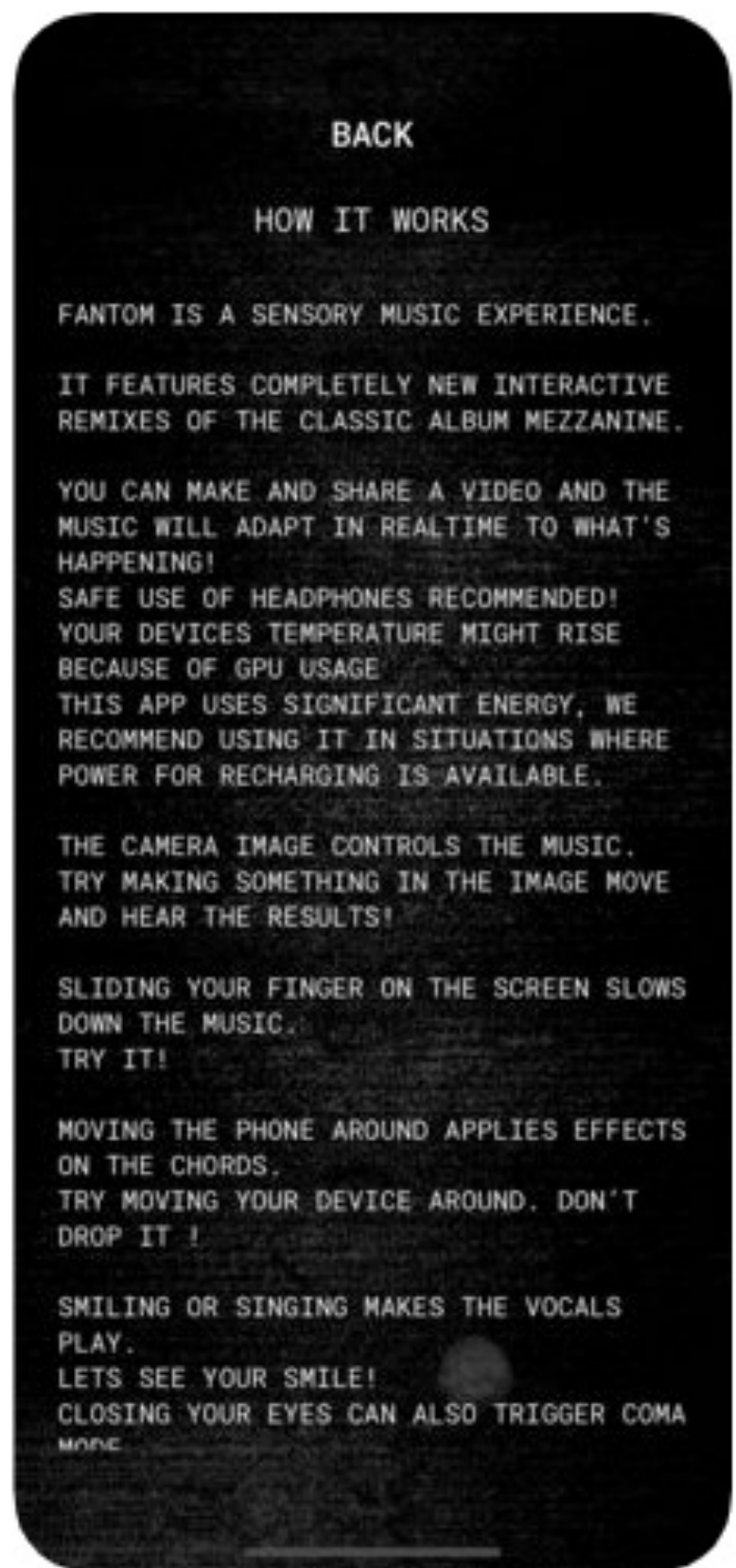
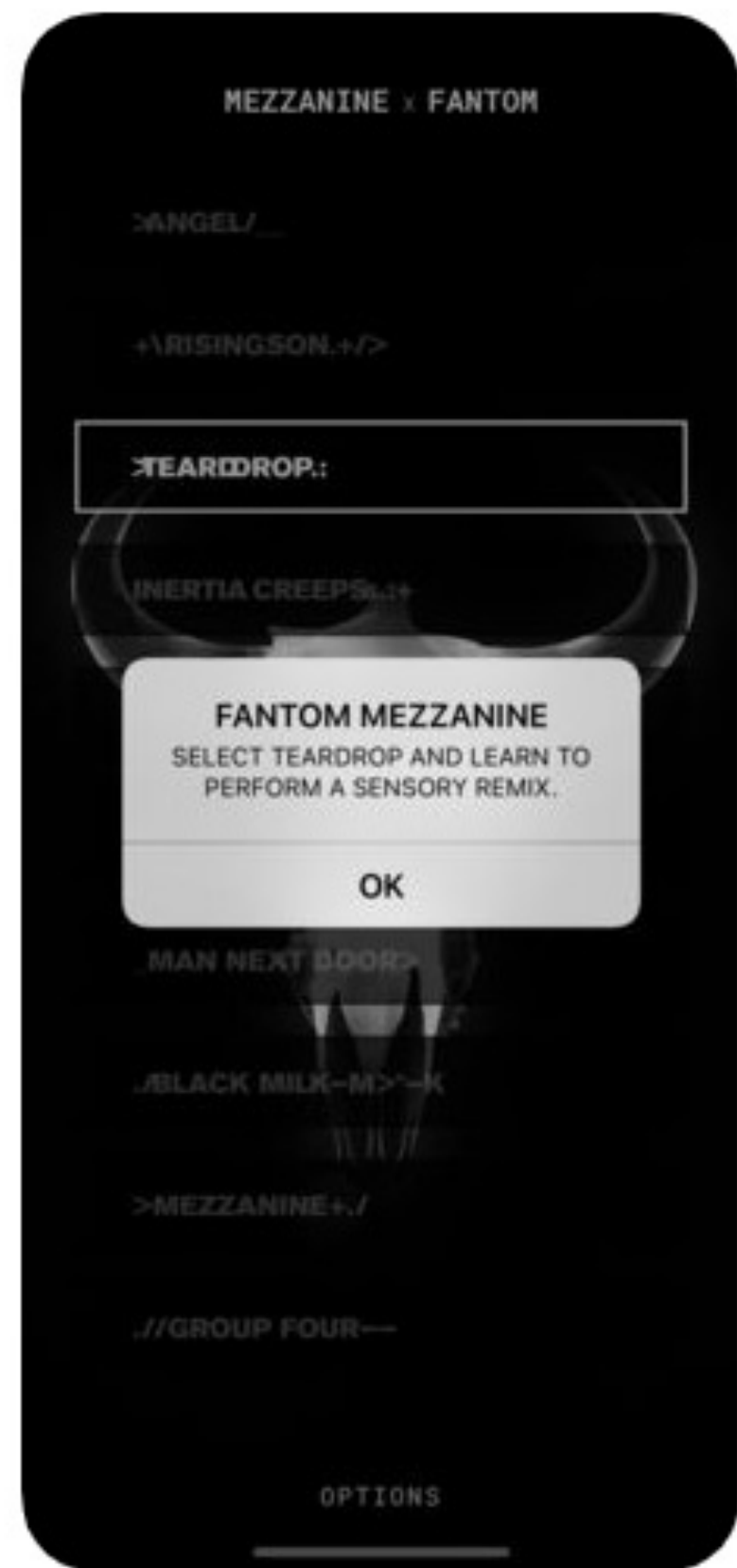
<https://vimeo.com/111858776>

<https://keirvine.com/living-symphonies/>



YouTube Smash Up

<https://pkmital.com/home/works/youtube-smash-up/>



Fantom

<https://www.thirdspace.co/fantom>



Make Music and Art Using Machine Learning

[Read the Blog](#)

[Try the Demos](#)

WHAT IS MAGENTA?

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

Magenta

<https://magenta.tensorflow.org/>

MusicVAE: Creating a palette for musical scores with machine learning.

Mar 15, 2018

Adam Roberts  adarob  ada_rob

Jesse Engel  jesseengel  jesseengel

Colin Raffel  craffel  colinraffel

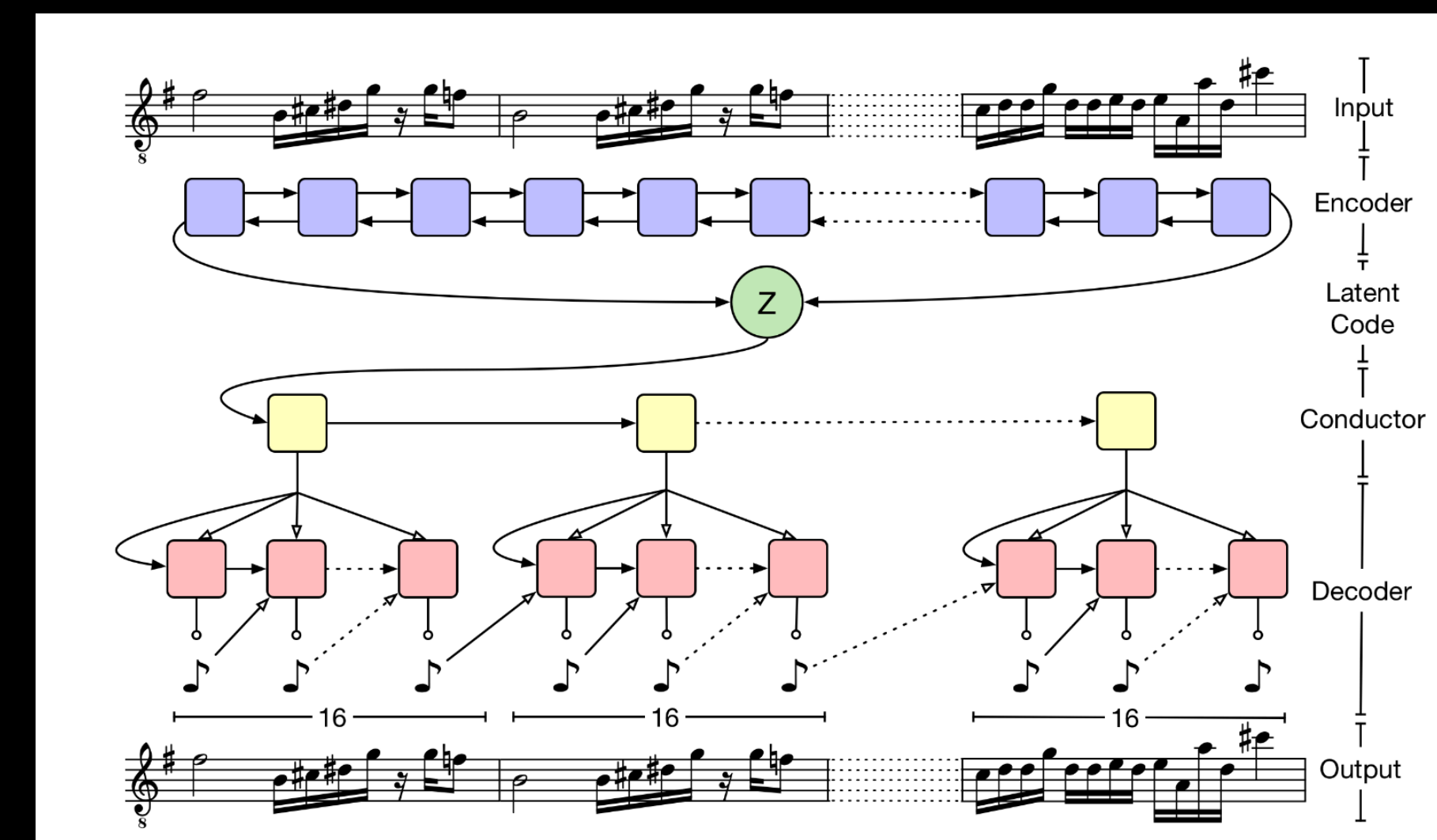
Ian Simon  iansimon  iansimon

Curtis Hawthorne  cghawthorne  fjord41

When a painter creates a work of art, she first blends and explores color options on an artist's palette before applying them to the canvas. This process is a creative act in its own right and has a profound effect on the final work.

Musicians and composers have mostly lacked a similar device for exploring and mixing musical ideas, but we are hoping to change that. Below we introduce **MusicVAE**, a machine learning model that lets us create palettes for blending and exploring musical scores.

As an example, listen to this gradual blending of 2 different melodies, A and B. We'll explain how this morph was achieved throughout the post.



Music VAE

<https://magenta.tensorflow.org/music-vae>



Dadabots
<https://dadabots.com/>



34.5TB

WDCH Dreams

<https://pkmital.com/home/works/walt-disney-concert-hall/>

MuseNet

OpenAI MuseNet

<https://openai.com/research/musenet>

Interacting with Jukebox

File Edit View Insert Runtime Tools Help

+ Code + Text Copy to Drive

Connect GPU High-RAM

Select mode

Run one of these cells to select the desired mode.

```
# The default mode of operation.
# Creates songs based on artist and genre conditioning.
mode = 'ancestral'
codes_file=None
audio_file=None
prompt_length_in_seconds=None

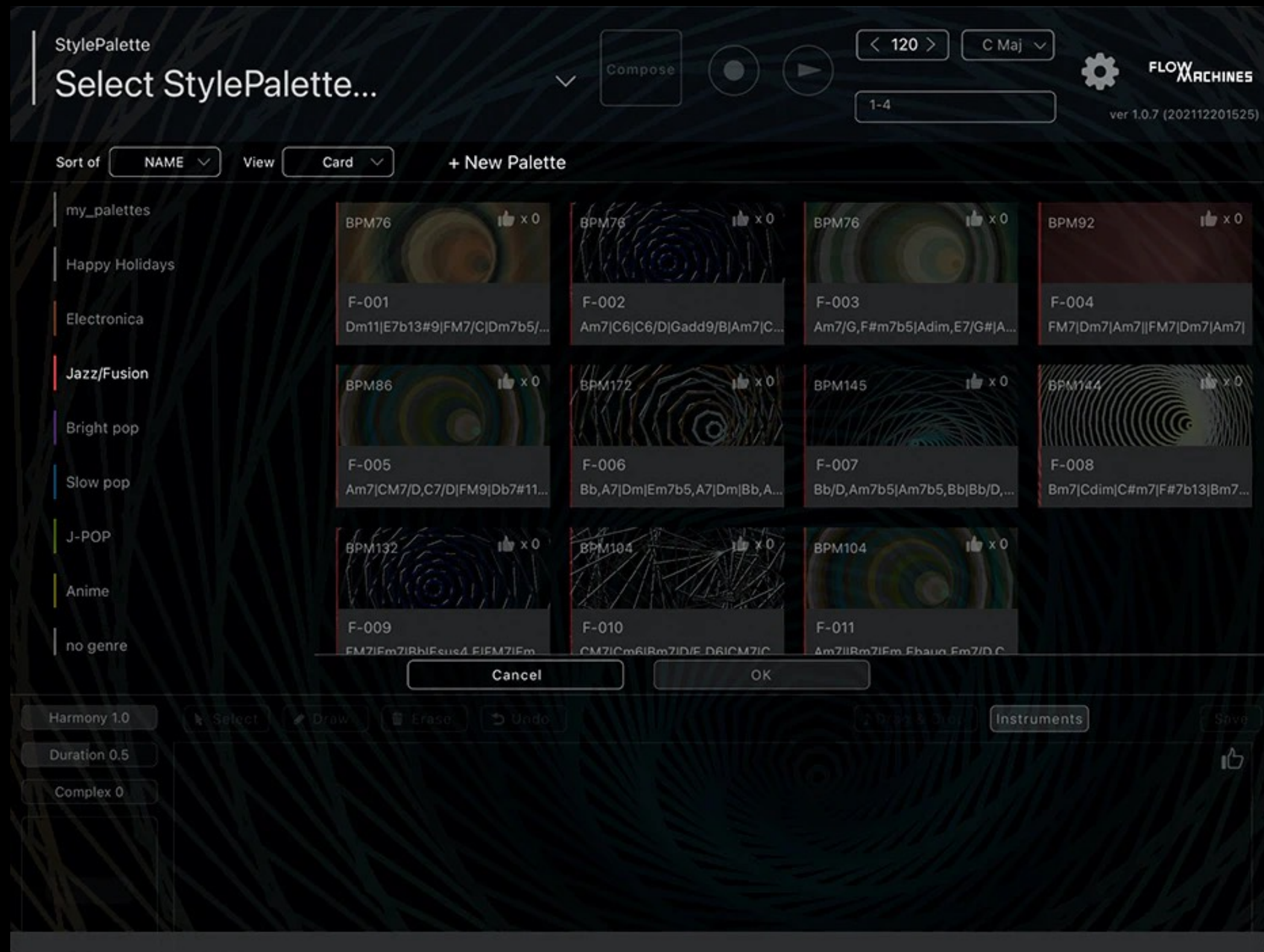
[ ] # Prime song creation using an arbitrary audio sample.
mode = 'primed'
codes_file=None
# Specify an audio file here.
audio_file = '/content/gdrive/My Drive/primer.wav'
# Specify how many seconds of audio to prime on.
prompt_length_in_seconds=12

Run this cell to automatically resume from the latest checkpoint file, but only if the checkpoint file exists. This will override the selected mode.
We will assume the existence of a checkpoint means generation is complete and it's time for upsampling to occur.

[ ] if os.path.exists(hps.name):
    # Identify the lowest level generated and continue from there.
    for level in [1, 2]:
        data = f"{hps.name}/level_{level}/data.pth.tar"
        if os.path.isfile(data):
            mode = 'upsample'
```

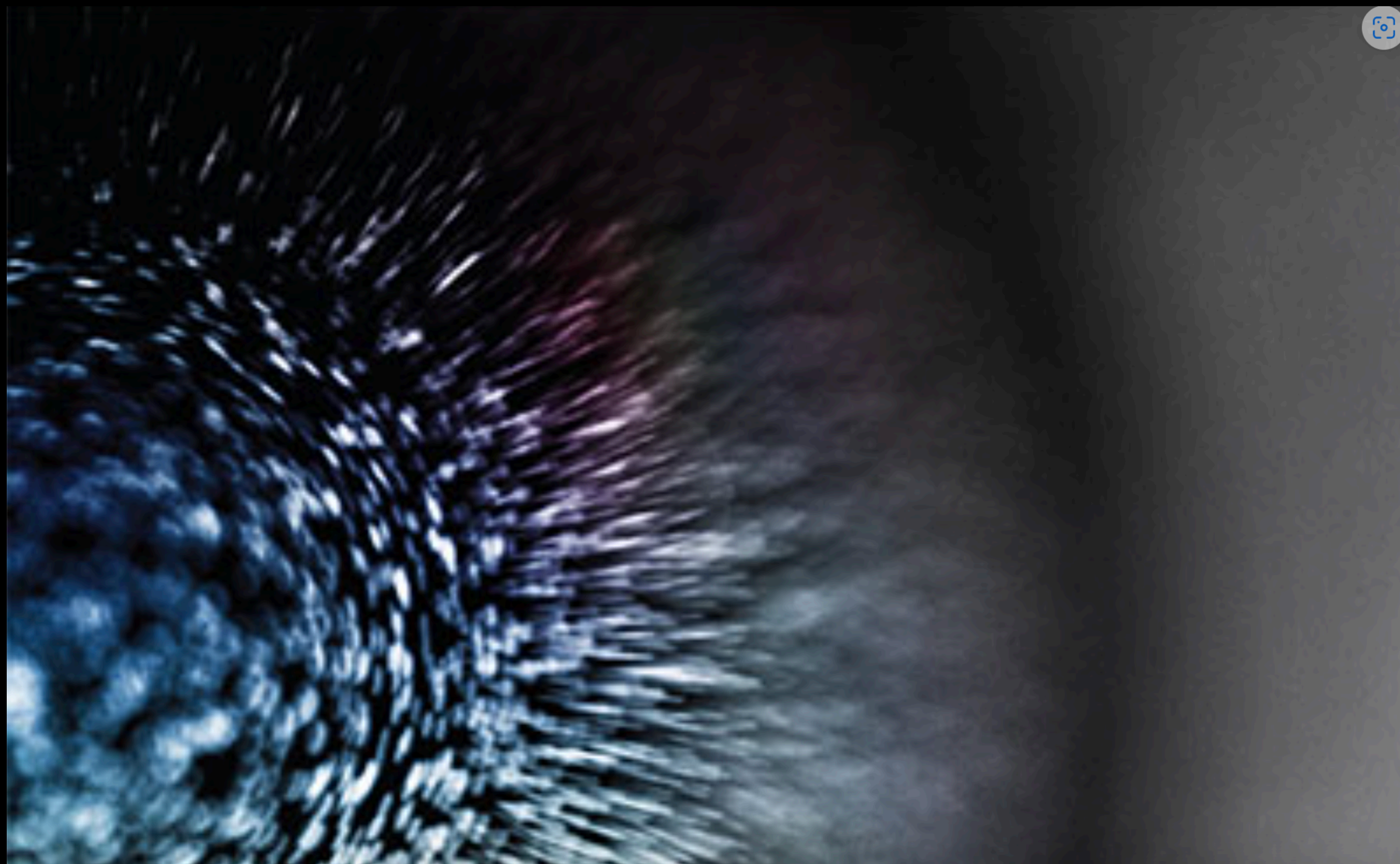
OpenAI Jukebox

[https://colab.research.google.com/github/SMarioMan/jukebox/blob/master/jukebox/Interacting with Jukebox.ipynb#scrollTo=VVOQ3egd65y](https://colab.research.google.com/github/SMarioMan/jukebox/blob/master/jukebox/Interacting%20with%20Jukebox.ipynb#scrollTo=VVOQ3egd65y)



Sony Flow Machines

<https://www.sony.com/en/SonyInfo/design/stories/flow-machines/>



October 8, 2019

EMPTYSET,
exploring
architecture,
algorithms &
the material
reality of
sound

Emptyset

<https://clotmag.com/interviews/emptyset-from-the-algorithmic-to-the-architectural-filling-the-spaces-between-our-ears>

DIY + UNUSUAL MUSIC SOFTWARE TECH

Watch Hexorcismos make AI performative and personal in this live show

Peter Kirn - July 26, 2023

In contrast to big data, automation, and bland prompts, artist Moisés Horta Valenzuela (Hexorcismos) is making machine learning into a live AV instrument using his own data sets and idiosyncratic, technoshamanistic approach. Check out this latest iteration of his live show Nahualtia Tlatzotzonalli (Shapeshifter Musician).

Moises Hexorcismos

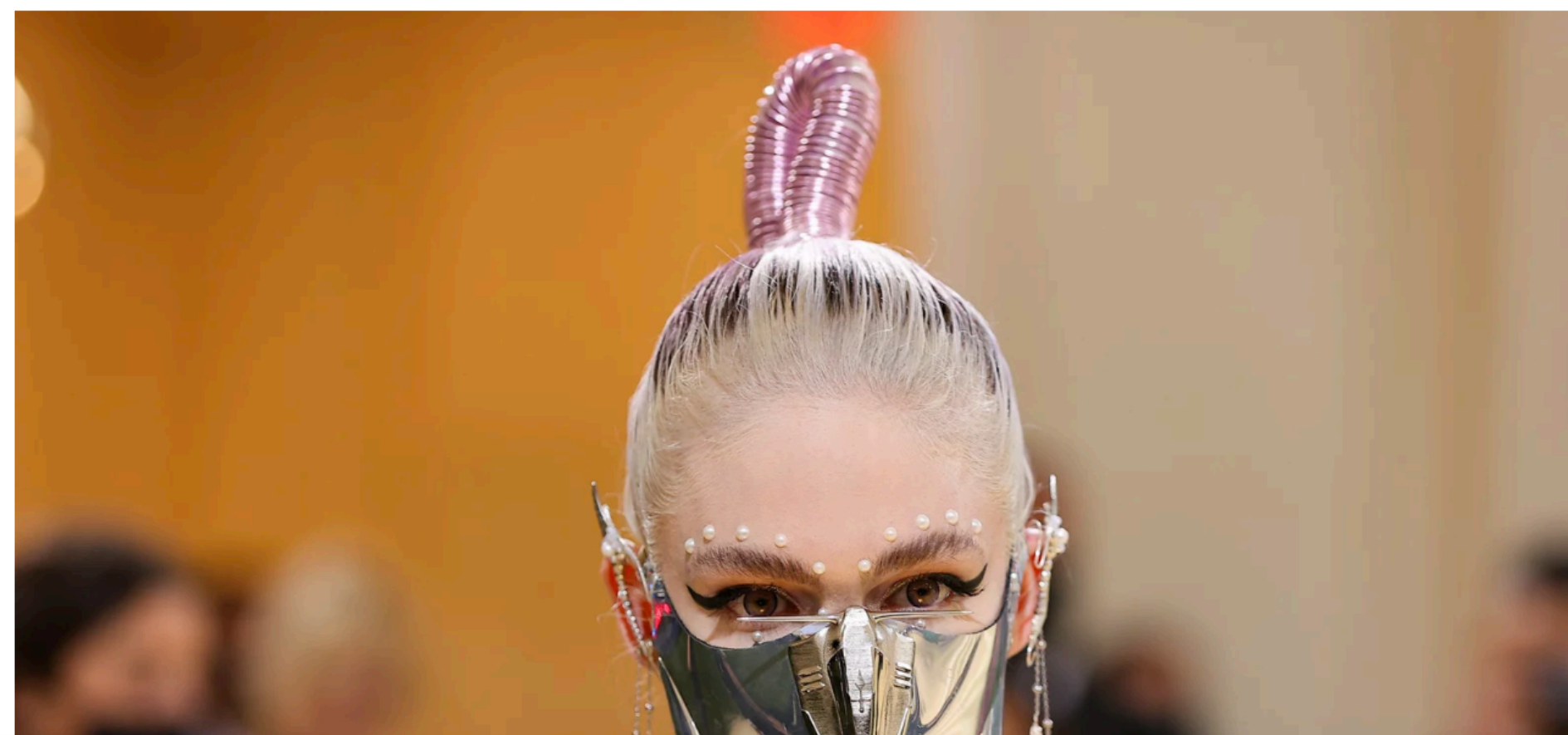
<https://cdm.link/2023/07/watch-hexorcismos-ai-live/>

MUSIC

Grimes invites fans to make songs with an AI-generated version of her voice

April 24, 2023 · 7:21 PM ET

 Vanessa Romo



Grimes - AI generated music

<https://www.npr.org/2023/04/24/1171738670/grimes-ai-songs-voice>

LAB

GENERATE MUSIC, SOUND, FX, AND FOLEY

Environmental sound generation


AudioGen: Textually-Guided Audio Generation

With AudioGen, we demonstrated that we can train AI models to perform the task of text-to-audio generation. AudioGen is an auto-regressive transformer language model model conditioned on text: Given a textual description of an acoustic scene, the model can generate the environmental sound corresponding to the description with realistic recording conditions and complex scene context.

[→ Read the paper](#)

[→ Get started](#)

<https://audiocraft.metademolab.com/audiogen.html>

New: Generate on  Discord

Powering audio creation with generative AI

Generate an infinite world of sounds - samples, instruments, sound effects, or textures,
instantly. All from scratch.

Tractor engine humming

Upload a sample

GENERATE

<https://www.audiogen.co/>

Create music with AI.

Start generating music for free.
No credit card needed.

Try it out

Trance, Ibiza, Beach, Sun, 4 AM, Progressive, Synthesizer, 909, Dramatic Chords, Choir, Euphoric, Nostalgic, Dynamic, Flowing



▶ 0:00 ●

1:35

<https://stableaudio.com/>

Untitled Session

20 %

Share

A.I. tools

00:00 00:05.46 00:10 00:20 00:30 00:40 00:50

Split stems
Isolate individual tracks from any audio clip.

Key and beat match
Nudge clips into their perfect spot.

Text-to-audio
Generate an original sample from any text prompt.

Swap voices
Change one voice into another.

S scissors ... Edit

+

▶

Sign in

Sounds Studio

BETA

Sign up to start exploring the latest A.I. capabilities for music.

A modern music production platform, powered by machine learning

<https://sounds.studio/>

CREATE YOUR BEATS WITH THE POWER OF AI

Welcome to the AI Music Generation

Try it for free



Hi, I'm Neo. Need any help?

<https://soundraw.io/>

[RIFFUSION]

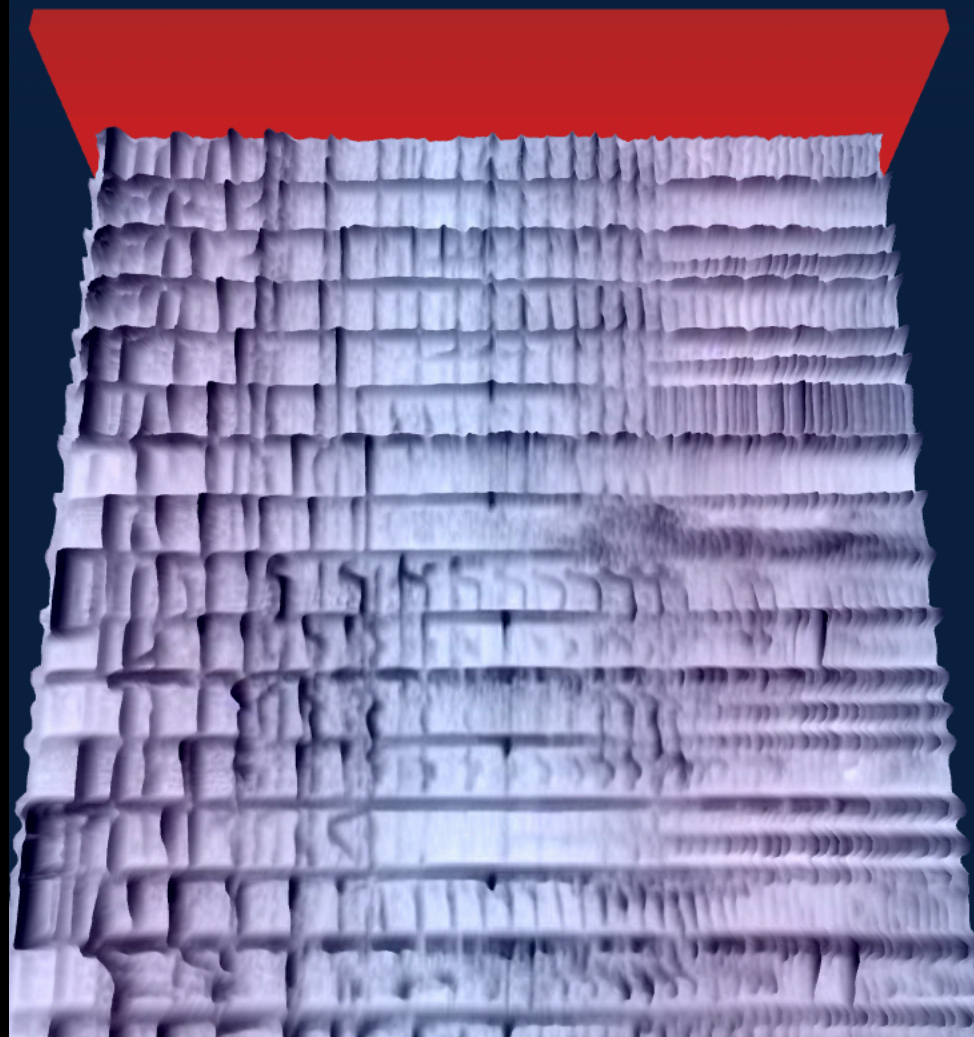
psychedelic nepalese trance

bubblegum eurodance

jamaican dancehall vocals

...

UP NEXT: Anything you want



<https://www.riffusion.com/>



Human \times AI Generative Music

For your video content, podcasts and apps

⚡ Generate a track now

Mubert Render

For content creators

Create a soundtrack that will fit your content's mood, duration and tempo. Instantly, easily, perfectly.

Make a track now

Mubert Studio

For artists

Earn money on tracks, samples and loops. Team up with AI to produce something incredible.

Contribute

Mubert Extension

For content creators

Create custom music within AE or Premiere with ease. Generate high-quality, original tracks hassle-free.

Download



<https://mubert.com/>

AudioLDM: Text-to-Audio Generation with Latent Diffusion Models

Haohe Liu^{*1}, Zehua Chen^{*2}, Yi Yuan¹, Xinhao Mei¹, Xubo Liu¹

Danilo Mandic², Wenwu Wang¹, Mark D. Plumley¹

¹CVSSP, University of Surrey, Guildford, UK

²Department of EEE, Imperial College London, London, UK


^{*}Equal Contribution

[\[Paper on ArXiv\]](#) [\[Code on GitHub\]](#) [\[Hugging Face Space\]](#)

Abstract

Text-to-audio (TTA) system has recently gained attention for its ability to synthesize general audio based on text descriptions. However, previous studies in TTA have limited generation quality with high computational costs. In this study, we propose AudioLDM, a TTA system that is built on a latent space to learn the continuous audio representations from contrastive language-audio pretraining (CLAP) latents. The


<https://audioldm.github.io/>

Musicfy  Artist Vocals Parody Vocals Instrumentals Music AI Sign In

Create AI Vocals

[Train your own AI Voice](#) →

Source



Select Audio File

or drag and drop mp3/wav file here. 15MB Limit

or

Start Recording

or

Youtube Link


Output

Select an Artist

Set Pitch

Remove Instrumental

Quality

All Artists Voices are Copyright Free 

<https://create.musicfy.lol/>

Make Music With AI Vocals

Epic synthetic singing and rapping vocals for creative agencies, musicians, and coders.

Get Started



<https://www.uberduck.ai/>

Sign up for AI Test Kitchen to create your own generative music

AI Test Kitchen is a place where people can experience and give feedback on some of Google's latest AI technologies. Our goal is to learn, improve, and innovate responsibly on AI together.

[Get started](#)



MusicLM is an experimental technology that allows you to generate your own synthetic music for inspiration. Certain queries that mention specific artists or include vocals will not be generated. If there is a problem with the generated audio, please leave us [feedback](#).

<https://aitestkitchen.withgoogle.com/experiments/music-lm>

Product Solutions Open Source Pricing

Search or jump to... Sign in Sign up

zhvng / open-musiclm Public

Notifications Fork 44 Star 366

Code Issues 7 Pull requests 2 Actions Projects Security Insights

main 5 branches 12 tags

Go to file Code

About

Implementation of MusicLM, a text to music model published by Google Research, with a few modifications.

arxiv.org/abs/2301.11325

artificial-intelligence transformer attention music-generation text-to-music

Readme MIT license Activity 366 stars 16 watching 44 forks Report repository

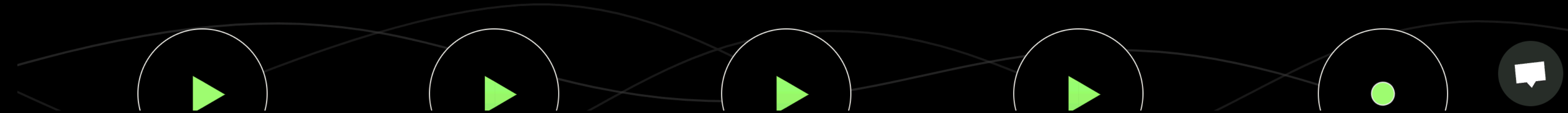
zhvng	update mert citation	8e2c6a8 on Jun 3	219 commits
configs	default configs		5 months ago
notebooks	make notebook smaller		5 months ago
open_musiclm	option to not use causal conv in transformer feed-forward		4 months ago
scripts	add option to generate continuation from input audio		5 months ago
.gitignore	update gitignore		5 months ago
LICENSE	Initial commit		8 months ago
MANIFEST.in	add manifest file		5 months ago
README.md	update mert citation		3 months ago
VERSION	add setup.py		5 months ago
clap.png	update readme		8 months ago
environment.yaml	change sklearn to scikit-learn #22		4 months ago

<https://github.com/zhvng/open-musiclm>

GENERATE VOICES

AI VOICE GENERATOR WITH REALISTIC TEXT TO SPEECH

Generate AI Voices, Indistinguishable from Humans



<https://play.ht/>

suno-ai / bark Public

Notifications Fork 3k Star 26.7k

Code Issues 115 Pull requests 9 Discussions Actions Projects Security Insights

main 2 branches 0 tags

Go to file Code

kmfreyberg Update README.md ...	61710e5 5 days ago	🕒 80 commits
📁 bark	Merge pull request #364 from no2chem/flash_attention	2 months ago
📁 notebooks	update memory profiling notebook	5 months ago
📄 .gitignore	Bark v1.5 (#226)	5 months ago
📄 LICENSE	Bark v1.5 (#226)	5 months ago
📄 README.md	Update README.md	5 days ago
📄 model-card.md	small updates	5 months ago
📄 pyproject.toml	Pin huggingface-hub dep	4 months ago
📄 setup.py	first commit	5 months ago

☰ README.md

Notice: Bark is Suno's open-source text-to-speech+ model. If you are looking for our new text-to-music

About

🗨️ Text-Prompted Generative Audio Model

📖 Readme

📄 MIT license

📈 Activity

☆ 26.7k stars

👁️ 276 watching

🍴 3k forks

Report repository

Used by 12



Contributors 16

<https://github.com/suno-ai/bark>

Coqui + Bark Voice Cloning

Mimic any voice character in less than 2 minutes with this [Coqui TTS + Bark](#) demo !

Upload a clean 20 seconds WAV file of the vocal persona you want to mimic,
type your text-to-speech prompt and hit submit !

[Duplicate this Space](#)

Text to speech prompt

One or two sentences at a time is better* (max: 10)

Hello friend! How are you today?

Waveform video



File upload

Microphone

Voices Characters

WAV voice to clone

Drop Audio Here

- or -

Click to Upload

Character Name

Name that voice character

description

How would you describe that voice ?

Clean sample ?

[Share with Community](#)

<https://huggingface.co/spaces/fffiloni/instant-TTS-Bark-cloning>

Plachtaa / VALL-E-X Public

Notifications Fork 452 Star 5.3k

Code Issues 26 Pull requests 2 Actions Projects Security Insights

master 1 branch 0 tags Go to file Code

Plachtaa Merge remote-tracking branch 'origin/master' 4d4964e 5 days ago 126 commits

customs	Update README.md	last month
data	updated requirements	2 weeks ago
images	update description	last month
models	Added batch decoding	5 days ago
modules	Fix typo in scaling.py	3 weeks ago
nltk_data/tokenizers/punkt	Provide offline NLTK resource files	3 weeks ago
presets	Replaced codec with vocos decoder	2 weeks ago
prompts	Updated UI	3 weeks ago
utils	fix text encoding	2 weeks ago
LICENSE	Create LICENSE	last month
README-ZH.md	Merge remote-tracking branch 'origin/master'	5 days ago

About

An open source implementation of Microsoft's VALL-E X zero-shot TTS model. Demo is available in <https://plachtaa.github.io>

text-to-speech tts gpt transformer-architecture emotional-speech voice-clone vall-e

Readme MIT license Activity 5.3k stars 62 watching 452 forks Report repository

Releases

<https://github.com/Plachtaa/VALL-E-X>

#1 AI VIDEO GENERATION PLATFORM

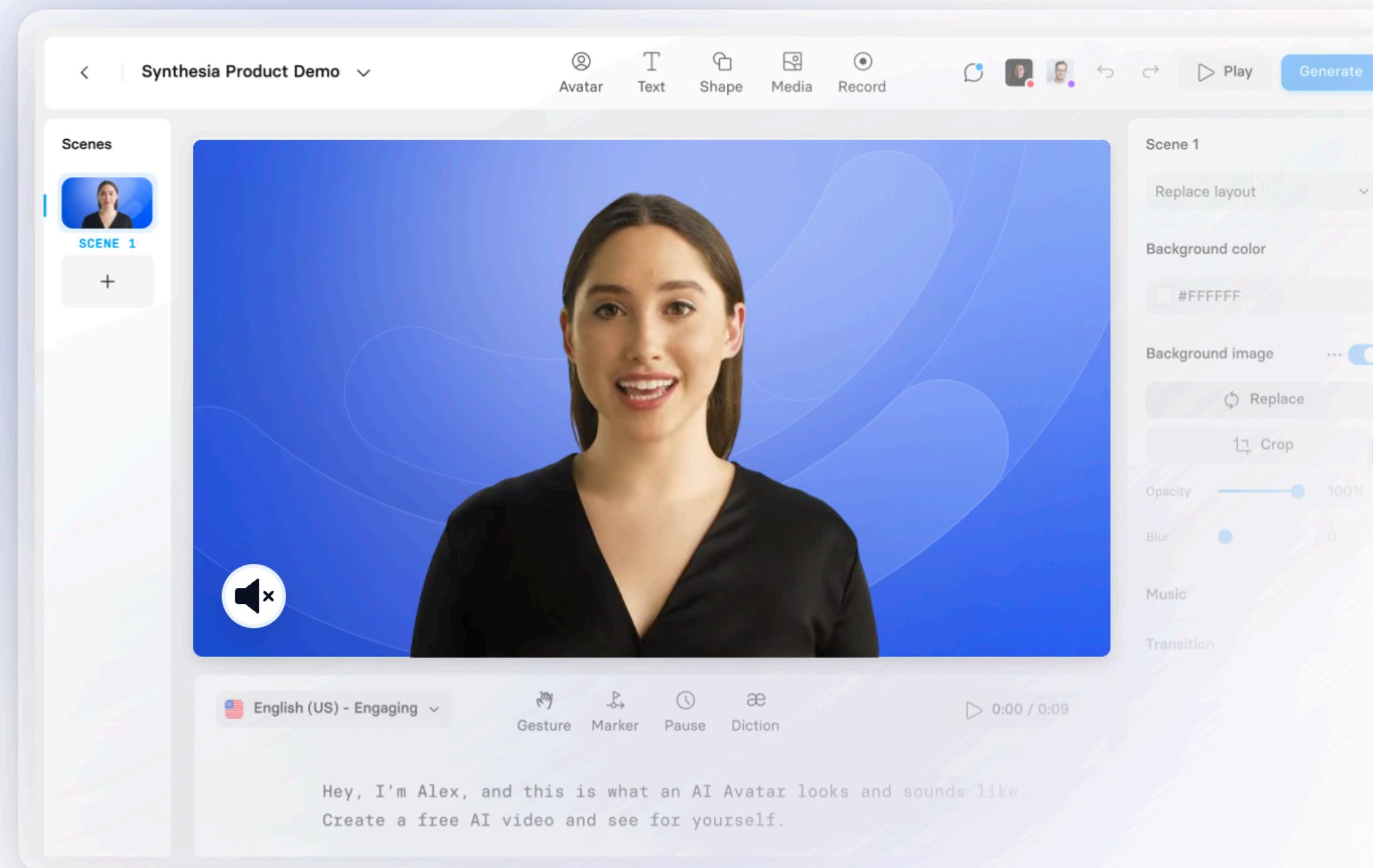
Turn your text into videos in minutes

- Get natural sounding AI voices in 120+ languages
- Make your videos more engaging with 140+ AI Avatars
- Edit as simply as a slide-deck, no experience required

Create a free AI video

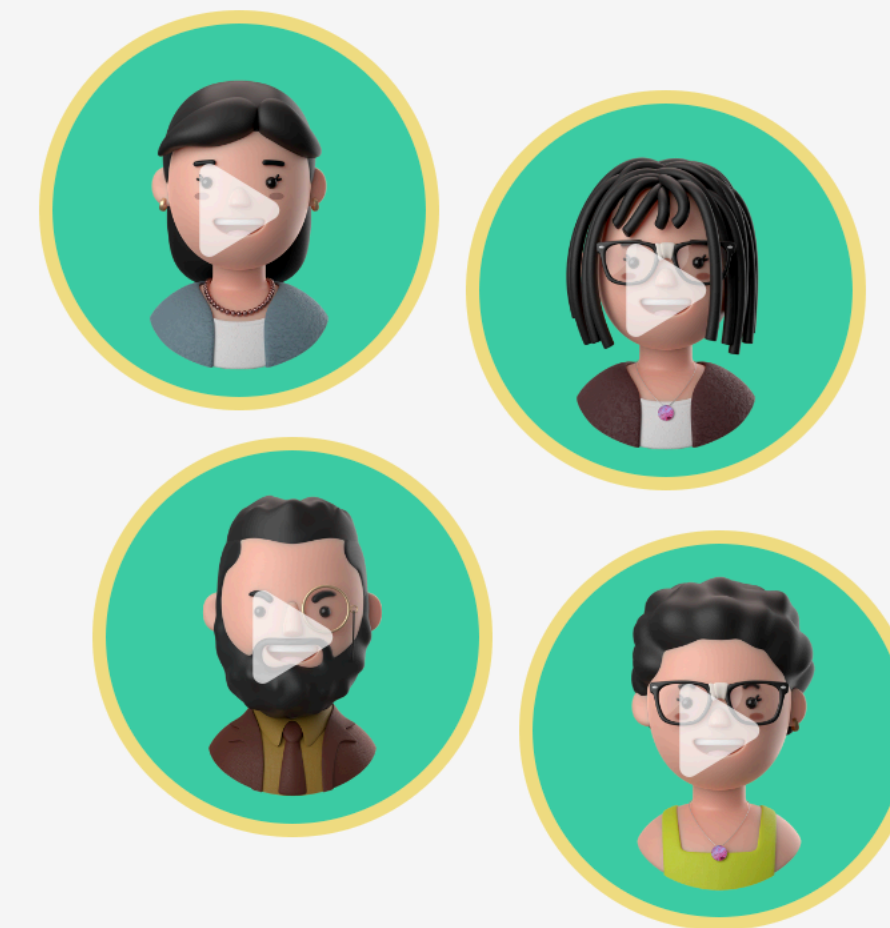
▶ Watch product tour

No credit card required.



<https://www.synthesia.io/>

Enterprise Ready Generative Voice AI Toolkit



✔ Text-to-Speech ✔ Speech-to-Speech ✔ Neural Audio Editing ✔ Language Dubbing

Resemble's AI voice generator lets you create realistic human-like voiceovers in seconds.

<https://www.resemble.ai/>

Spaces | justest/wav2lip | like 2 | Running

App | Files | Community

Wav2lip

Gradio demo for Wav2lip: Accurately Lip-syncing Videos In The Wild. To use it, simply upload your image and audio file, or click one of the examples to load them. Read more at the links below. Please trim audio file to maximum of 3-4 seconds

image

Drop Image Here
- or -
Click to Upload

audio

Record

Clear

Submit

Screenshot

<https://huggingface.co/spaces/justest/wav2lip>

Coqui Studio XTTS: Personalities Saved breannabrowning@gmail.com B

My Projects > XTTS: Personalities

Meet the Gang + Scene Description ↶ ↷ Invite Export

⋮ + New performance Paraphrase Speech Rate Language: English Duplicate line Delete line Save as take

CHARACTER
z Zofija Kendrick 5 | 250

⋮ + New performance Paraphrase Speech Rate Language: English Duplicate line Delete line Save as take

Wow.

CHARACTER
x Xavier Hayasaka 73 | 250

⋮ + New performance Paraphrase Speech Rate Language: English Duplicate line Delete line Save as take

If you were me, then I'd be you! No matter who you are you can't stop me!

00:00.000 - 00:20.255

Characters

- Aaron Dresch...
- Alexandra Hi...

<https://coqui.ai/>

HeyGen Create Video


Sign In For Free Credits!

- Home
- Template
- Avatar
- Voice
- Video
- Asset

Webinar Live Soon >

- Pricing
- Labs
- Tutorial
- News

Create Avatar



Instant Avatar
Get your digital twin in minutes!






Photo Avatar
Animate your photo with text

Create Video



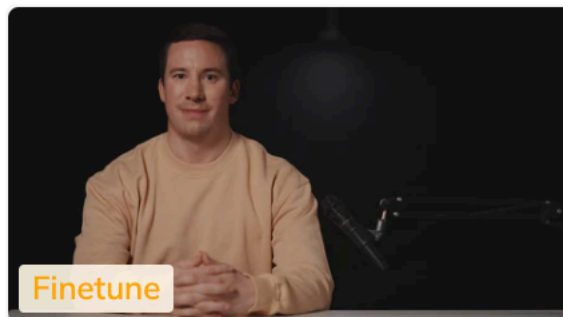
Template
Start from a template



AI Script
Start from AI script Generation

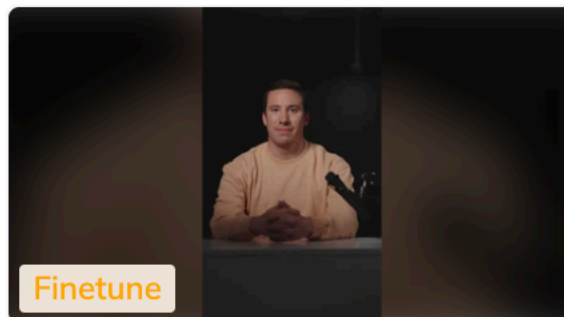
Your first digital twin for free! Ready in 5 minutes!

[Create Instant Avatar](#)



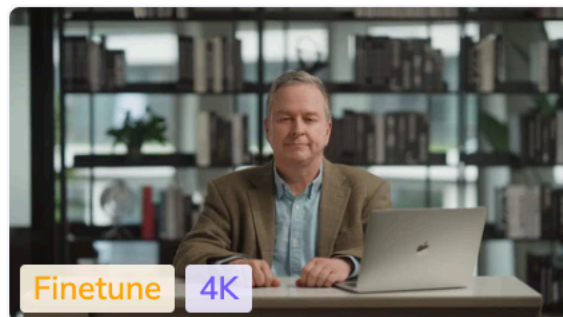
Finetune

Blake




Finetune

Blake




Finetune 4K

Edward



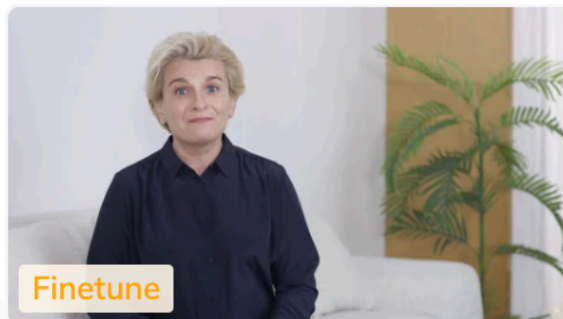
Finetune

Leah



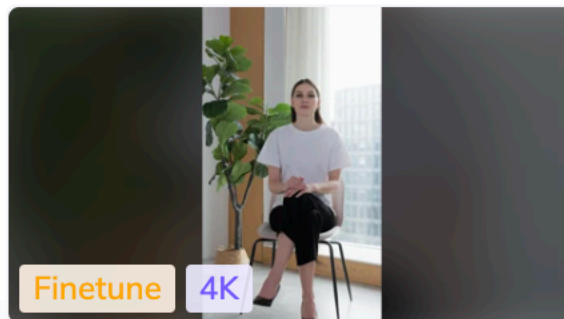
Finetune 4K

Matthew



Finetune

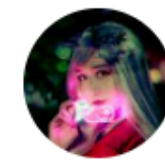
Teva



Finetune 4K

Vera

<https://app.heygen.com/guest/home>



Aimi Sekiguchi 愛美
@sekiguchiaimi



AI変換ビデオビフォーアフター 🇯🇵 → 🇬🇧

AI generated video before and after.

#HeyGen #aigenerated

app.heygen.com/guest?sid=rewa...

Translate post



1:05 AM · Sep 24, 2023 · 955.8K Views

<https://twitter.com/sekiguchiaimi/status/1705855926148984915?s=46&t=MHjSXuuPGHmWLD4It3ZkFA>

Other models for voice generation

- <https://github.com/MycroftAI/mimic3>
- https://openreview.net/forum?id=__czv_gqDQt
- <https://github.com/heatz123/naturalspeech>
- <https://github.com/legekka/GanyuTTS>
- <https://github.com/rendchevi/nix-tts>
- <https://github.com/mush42/piper-nvda>

SEGMENT AUDIO

Hugging Face Search models, datasets, users... Models Datasets Spaces Docs Solutions Pricing Log In Sign Up

Spaces: Matthijs/whisper_word_timestamps like 98 Runtime error on T4 App Files Community 4

main whisper_word_timestamps / app.py

Matthijs Hollemans add language selector d2d20b7 3 months ago

</> raw history blame contribute delete No virus 9.42 kB

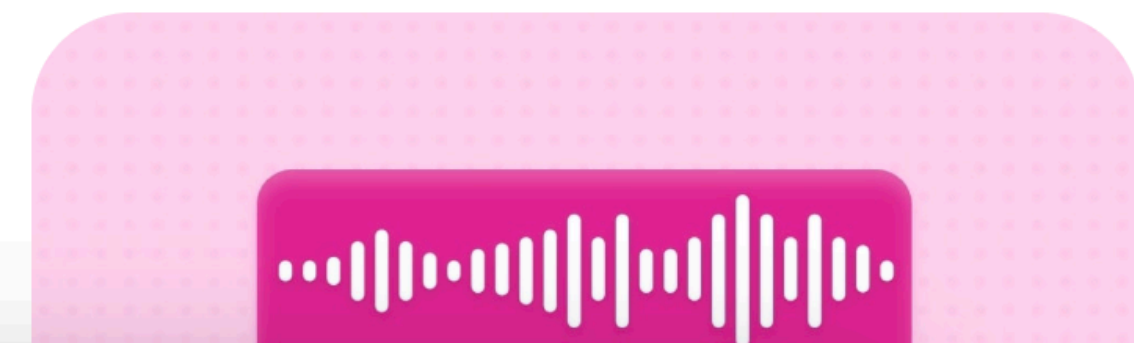
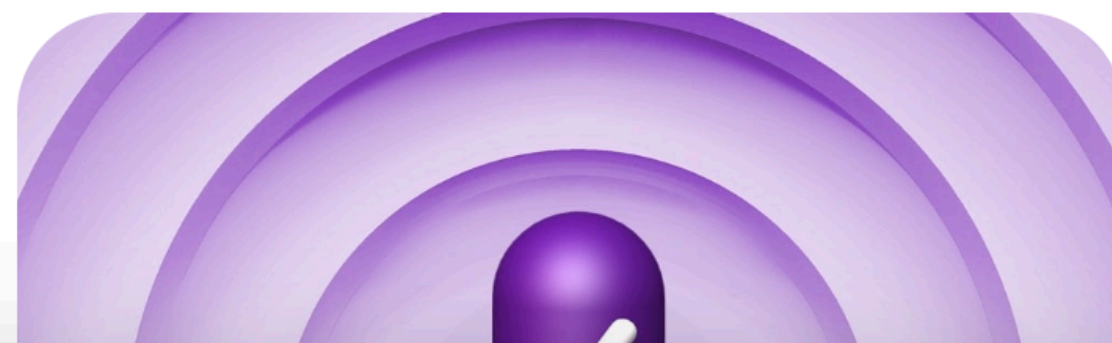
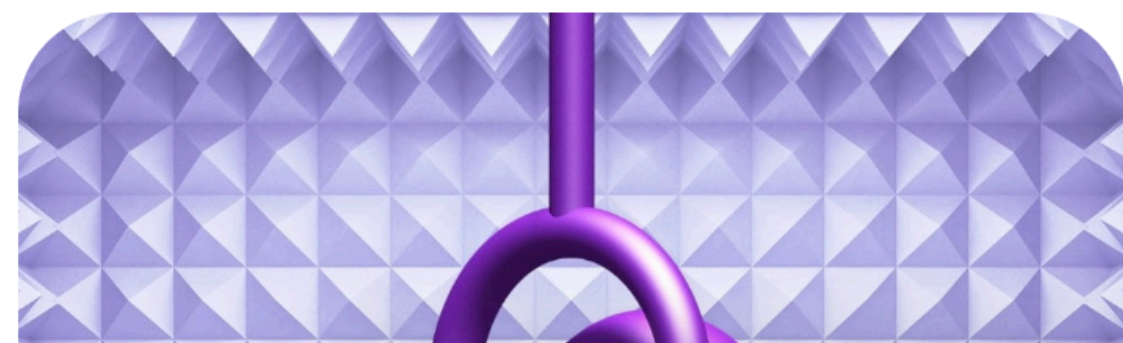
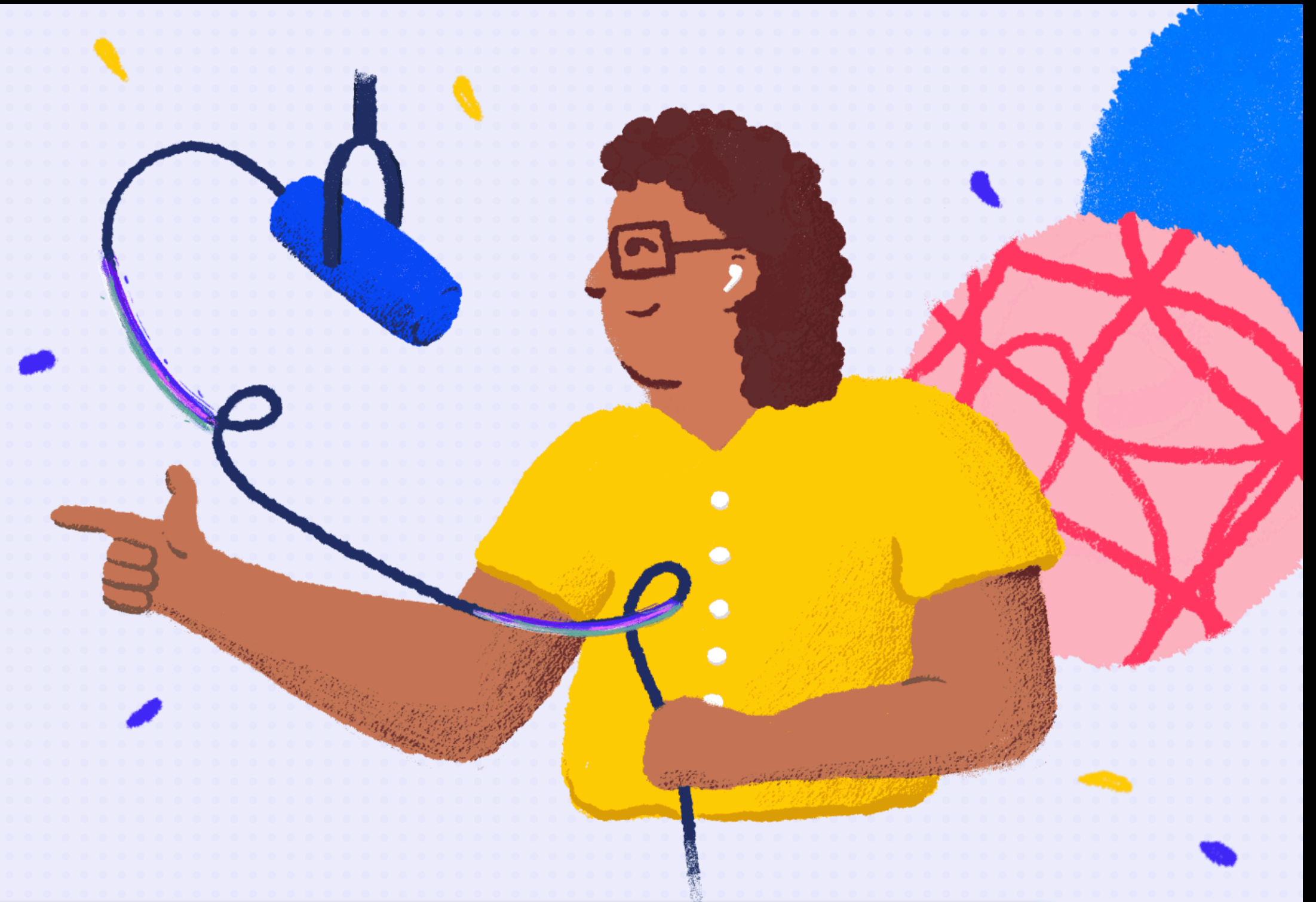
```
1 import gradio as gr
2 import librosa
3 import numpy as np
4 import moviepy.editor as mpy
5 import torch
6
7 from PIL import Image, ImageDraw, ImageFont
8 from transformers import pipeline
9
10
11 # checkpoint = "openai/whisper-tiny"
12 # checkpoint = "openai/whisper-base"
13 checkpoint = "openai/whisper-small"
14
15 # We need to set alignment_heads on the model's generation_config (at least
16 # until the models have been updated on the hub).
17 # If you're going to use a different version of whisper, see the following
18 # for which values to use for alignment_heads:
19 # https://gist.github.com/hollance/42e32852f24243b748ae6bc1f985b13a
```

https://huggingface.co/spaces/Matthijs/whisper_word_timestamps/blob/main/app.py

CLEAN VOICES

AI-powered audio tools that elevate your voice

Create high-quality podcasts and voiceovers that sound professional with Adobe Podcast.



<https://podcast.adobe.com/>

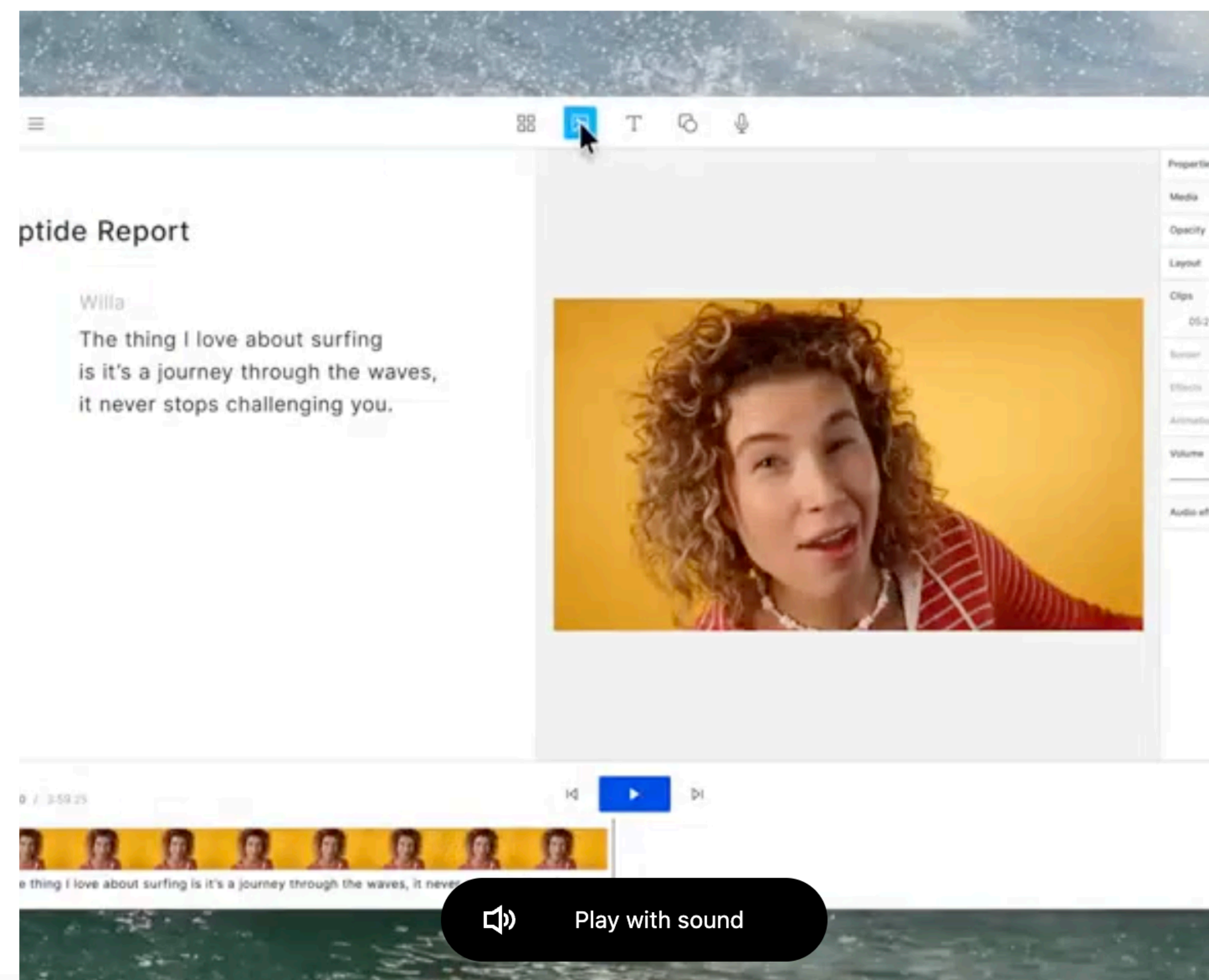


There's a new way to make video and podcasts. **A good way.**

Descript is the simple, powerful, and fun way to edit.

Get started for free →

See product tour



<https://www.descript.com/>

DEFAULT

HOW IT WORKS DONATE

Splitter AI

Split music into separated parts with AI-Powered algorithms

The screenshot displays the Splitter AI web interface. On the left is a dark sidebar with a menu icon at the top and several navigation buttons: Remover, Splitter (highlighted with a blue bar), Pitcher, Key BPM Finder, Cutter, Joiner, Support, and a UK flag. The main content area has a dark background with the title 'Splitter AI' and a subtitle 'Split music into separated parts with AI-Powered algorithms'. Below this is a visualization of audio waveforms for 'Music', 'Vocal', 'Bass', and 'Drums', each with a corresponding volume slider. The 'Music' track is green, 'Vocal' is blue, 'Bass' is yellow, and 'Drums' is red. The interface also includes links for 'HOW IT WORKS' and 'DONATE' at the top right.

<https://vocalremover.org/splitter-ai>

Music & Voice Separation

MVSEP performs separation of audio on voice and music parts

Drag & Drop to Upload File

OR


[Browse File](#)

Separation type

Model type

Output encoding

Put on [demo page](#)

I'm not a robot 
reCAPTCHA
[Privacy](#) - [Terms](#)

[Separate](#)

Unprocessed files in queue: 32. Currently processed with GPU: 6

<https://mvsep.com/en>

ARCHIVE

Archival resources / audio generation models

- audio style transfer: https://colab.research.google.com/drive/18_DckMGM-fsSthlqZI9sOAIY6l3plDLD#scrollTo=aEH0RfPbasnP
- voice style transfer: <https://colab.research.google.com/drive/1tacJzTEJW8Y42kBu-NiYQdVHuOOYbNZ6>
- voice style transfer with unpaired collections: <https://colab.research.google.com/drive/16t0TVX1D6y941d4uFL32e2eNxeDmB-Xi#scrollTo=V00rptcdKSbq>
- speech to text to speech: <https://colab.research.google.com/drive/100X8F1uaul6rk1BRbpDZxtLWRle7CF7r#scrollTo=p-Bi7A11-cll>
- speech to text from youtube: <https://colab.research.google.com/github/tugstugi/dl-colab-notebooks/blob/master/notebooks/MozillaDeepSpeech.ipynb#scrollTo=fZwM0GtmA7mX>
- source separation (complex to stems, including voice): <https://colab.research.google.com/github/deezer/spleeter/blob/master/spleeter.ipynb#scrollTo=f8Brdfh6mzEz>
- source separation (complex to stems, including voice): <https://colab.research.google.com/drive/1mijF0zGWxN-KaxTnd0q6hayAlrID5fEQ>
- openai jukebox: https://colab.research.google.com/github/SMarioMan/jukebox/blob/master/jukebox/Interacting_with_Jukebox.ipynb#scrollTo=65aR2OZxmzfzq

Further approaches

- **AIVA:** An AI-powered software that can compose classical music in a range of styles
- **Amper Music:** An AI-powered tool that allows users to create custom music tracks by selecting parameters like mood, style, and pace
- **Sonix:** An AI transcription tool that automatically transcribes audio recordings into written text
- **Descript:** A tool that uses AI to produce accurate transcripts that you can edit by editing the text, which in turn edits the audio

Resources

- Cycling 74
- Pure Data
- Bob Sturm
- Magenta
- GRUV
- Nick Collins
- Ali Momeni Max MSP
- Rebecca Fiebrink Wekinator and Kadenze Course
- NIME
- Neurips Creativity Workshop
- ProcGen
- ISEA
- MAH

HOMework

Homework

- No lab / student presentations due to class cancellation Wednesday October 25
- Students should explore audio generation tools independently, considering their application in final projects