

feature.AI

By Andrew, Jason, Sahil, Aarav

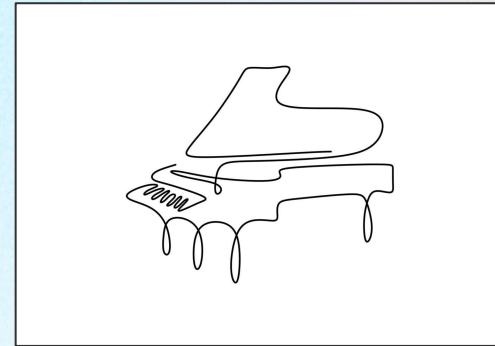
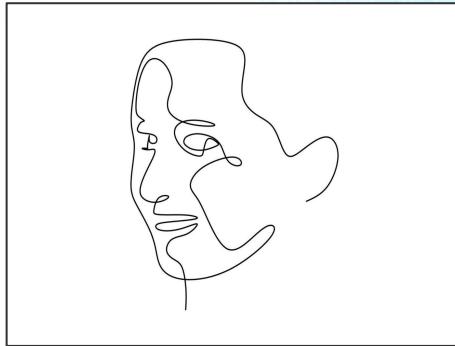
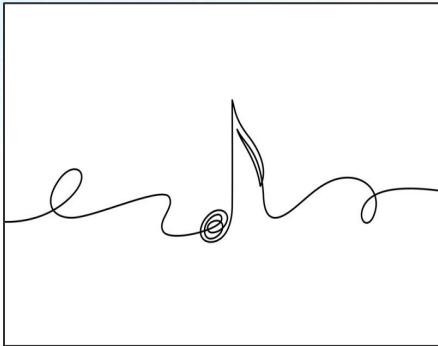


TABLE OF CONTENTS

01

context

Background and context

02

process

Step-by-step walkthrough on
our project

03

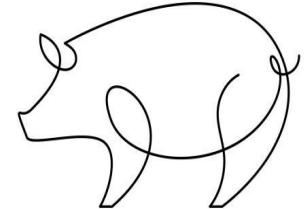
conclusion

Findings, challenges and
conclusions

04

demo

AI music demo with newly
generated music



01

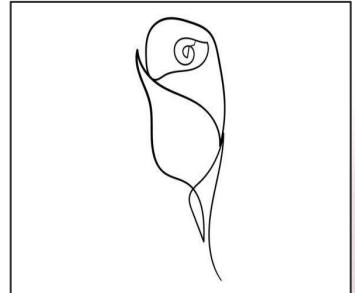
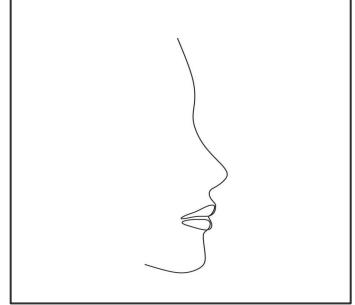
context

CONTEXT summary CHALLENGES

Create an AI music snippet generator from scratch

- Isolate lyrics from melody
- Generate AI celebrity voice and lyrics
- Transpose celebrity voice onto snippet, then add back into original

This is our 3-step process for generating new artist features on existing music with AI. We attempt to add lyrics to songs with the help of AI, Retrieval-based Voice Conversion, Fourier Transforms, and Latent Diffusion.

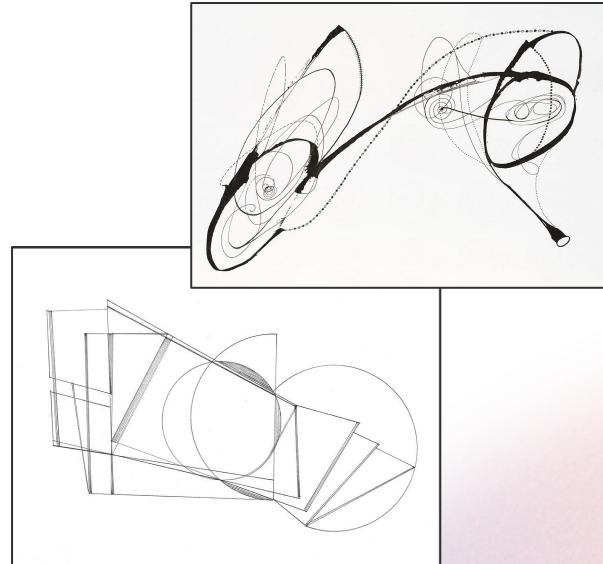


CONTEXT summary CHALLENGES

Difficult Challenges we must overcome:

- Music is difficult to represent as data, has inherent potential for human interpretation
- Must ensure that AI-generated lyrics matched both the mood and structure of the original track

We had to figure out a way to quantify music. Additionally, music is an art – it can be interpreted differently. It required a unique approach compared to a conventional data analysis project.



02

Process

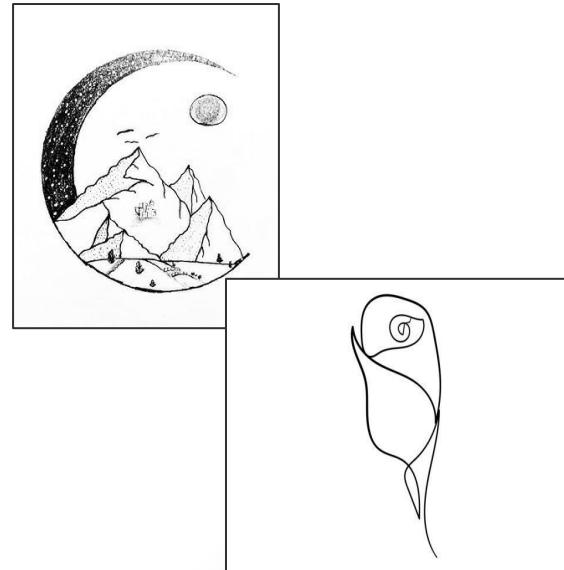
Process

ISOLATE GENERATE TRANSPOSE

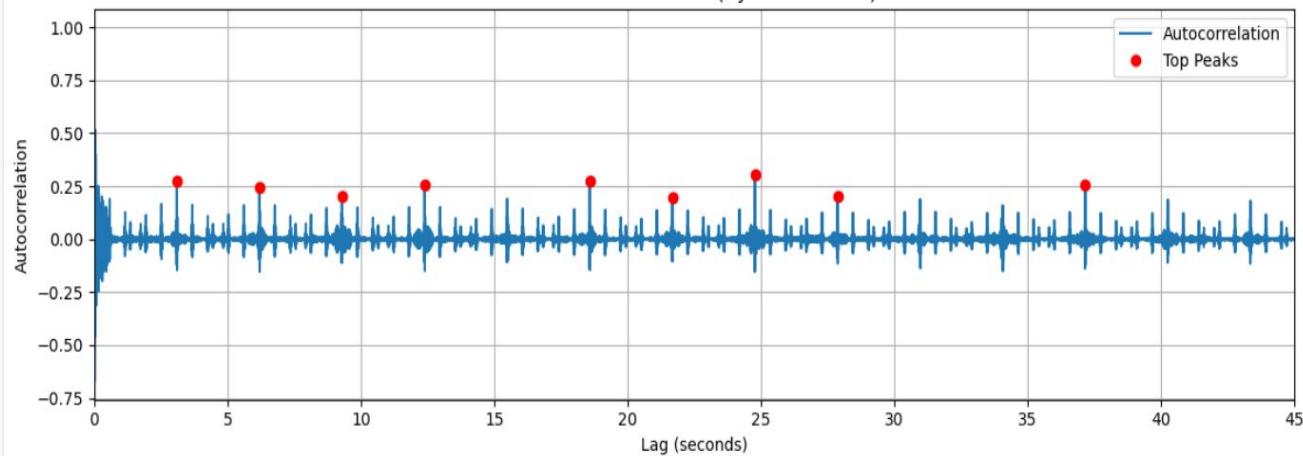
Isolate melody, drums, and bass from vocal elements

- Utilized Open-Unmix LSTM (Long-Short Term Memory) deep-learning model to estimate magnitude of components
- Pre-trained neural network with Short-Time Fourier Transform for element isolation
- Recursively iterated over sound for noise reduction and clarity

The end result was two distinct soundtracks – one melody, one vocal. We can now directly work with just the melody; it is a blank canvas for creative liberty.

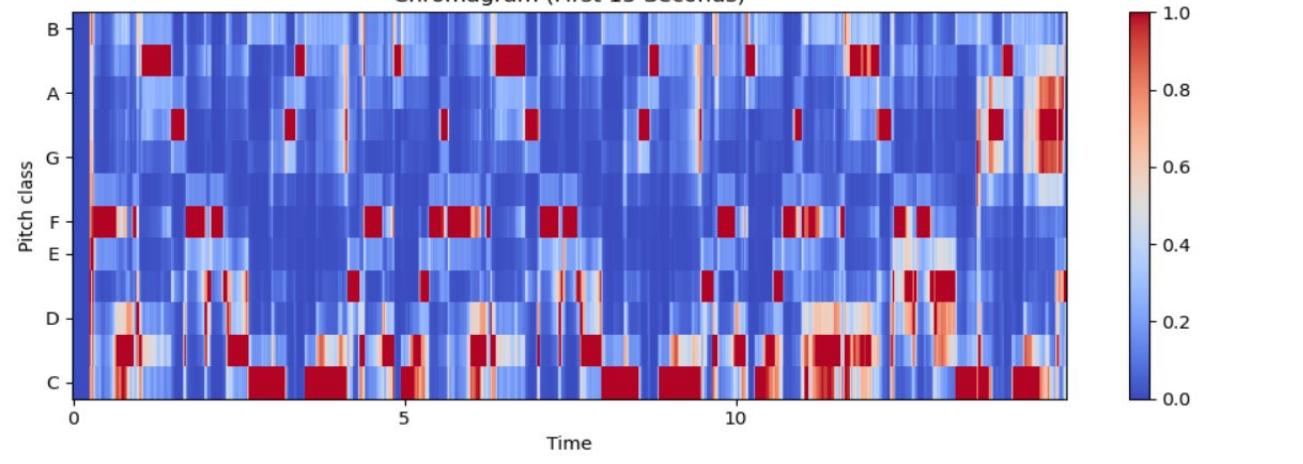


Autocorrelation Peaks (Cycle Detection)



- Algorithm to analyze cyclical patterns in melody (motifs)
- Red dots are peak intensities in melody

Chromagram (First 15 Seconds)



- Short-Time Fourier Transform for pitch classification
- Analyzes intensity of note in 12-TET scale (C, D, E, F...) with respect to time

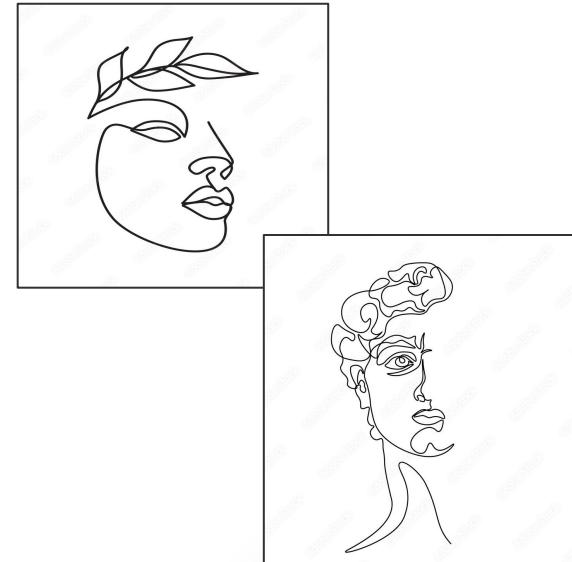
Process

ISOLATE GENERATE TRANPOSE

Generate our new lyrics

- Created AI-generated lyrics from DeepSeek API calls
- Developed a generic voice synthesis model to sing the lyrics
- Compiled new music using DiffRhythm from the HuggingFace repository

Our end result was an AI-generated snippet with unique melody *and* lyrics. However, we are not done yet. We must now make the lyrics sound like our celebrity of choice.



QuickWick	Delete REU Music (RVC) 350 Epoch/REU Music (RVC) 350 Epoch.zip	5d4b3c6
\$NOT (RVC) 400 Epoch		Upload 50 files
21 Savage (RVC) 1k Epoch		Upload 50 files
21 Savage 100k		Upload 12 files
21 Savage 50k		Upload 21 Savage 50k.zip
2Pac Tupac (RVC) 150 Epoch		Upload 93 files
2Pac Tupac 33k		Upload pac33k.zip
2Pac Tupac 50k		Upload 12 files
50 Cent (RVC) 1k Epoch		Upload 50 files

- A snapshot of the HuggingFace repository used to train voices
- Dataset differs in artist, training epochs, steps, etc

	File Name	Song Title	Artist	Youtube Link
0	Someone You Loved.mp3	Someone You Loved	Lewis Capaldi	www.youtube.com/watch?v=zABLEcsR5UE
1	Shape of You.mp3	Shape of You	Ed Sheeran	www.youtube.com/watch?v=JGwWNGJdvx8
2	Blinding Lights.mp3	Blinding Lights	The Weeknd	www.youtube.com/watch?v=fHI8X4OXluQ
3	As It Was.mp3	As It Was	Harry Styles	www.youtube.com/watch?v=H5v3kku4y6Q
4	Sunflower - Spider-Man	Sunflower - Spider-Man_ Into the Spider-Verse	Post Malone/Swae Lee	www.youtube.com/watch?v=ApXoWvfEYVU

- Additional dataset on music mp3s, artist names, song names, urls
- Used to source our data for song generation

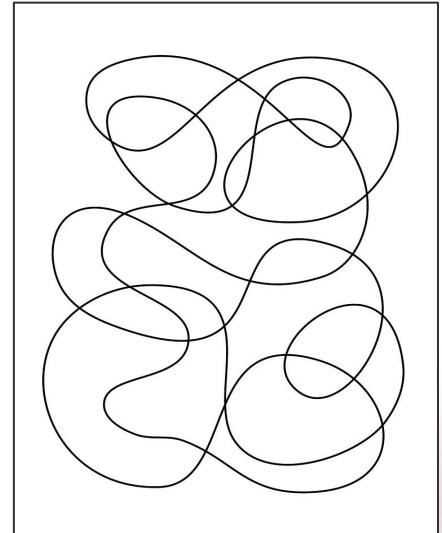
Process

ISOLATE Generate Transpose

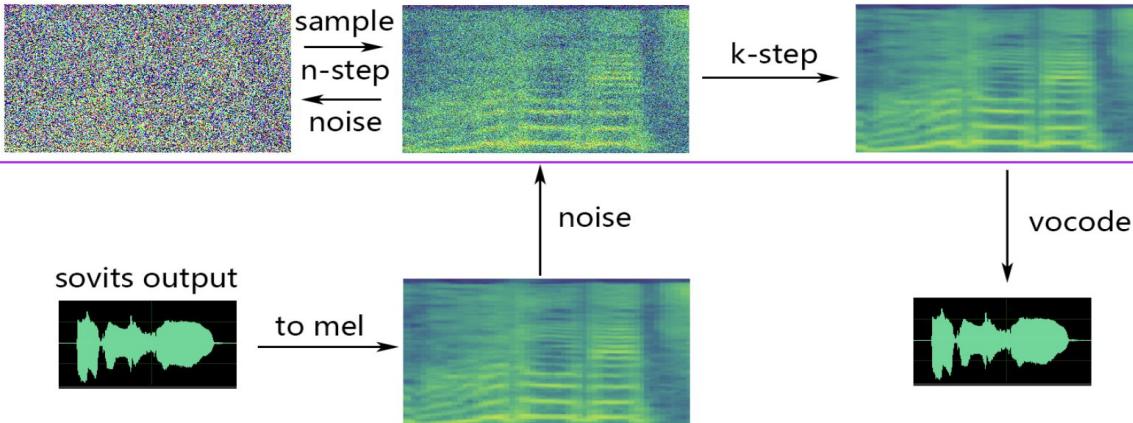
Employed Retrieval-based Voice Conversion (RVC) to transpose generic lyrics to celebrity soundalikes.

- Transforms raw data into a compact latent space using techniques like autoencoders, variational autoencoders (VAEs)
- Deep learning models predict song structure with artist projection
- Preserves semantics, intonation, and timbre of original song

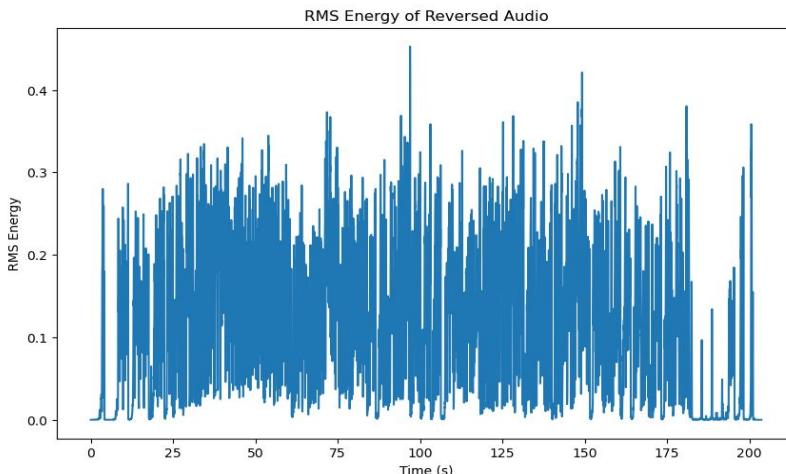
We have now converted our lyrical audio to sound like the chosen artist, instead of a generic Text-to-Speech. Our final step is to combine everything into one song



Diffusion model



- Music inputted is converted to spectrogram
- Controlled noise is added and removed over k-steps
- Vocode, or “convert” newly created spectrogram to song



- Graph of Root Mean Square (RMS) Energy of songs
- Measurement of a sound signal's energy and loudness over time
- Quantifiable estimate of its intensity

03

conclusion

conclusion

FINDINGS



We created original music with AI-generated lyrics and custom artist voices. Using Retrieval-based Voice Conversion (RVC), we transformed generic vocals to mimic celebrity voices while preserving the original meaning, tone, and style.



Given only 24 hours, we focused on building a functional pipeline. With more time, we'd improve lyric generation quality and train longer.

We are also considering sine wave manipulation and advanced signal processing techniques for smoother, higher-fidelity sound.

04

demo



Like I'm Gonna Lose you by Meghan Trainor
(f.t. Ariana Grande AI)



That's What I Like by Bruno Mars
(f.t. Michael Jackson AI)

THANK YOU