

# DADABOTS

## EXTENDED PROCESS DOCUMENT + ANNOTATION

AI Song Contest 2024

[\[SOUNDCLOUD\]](#)

[\[MUSIC VIDEO\]](#)

Music - CJ Carr & Zack Zukowski (DADABOTS)

Video - Mahdi Riahi (NOXBITE)



## **GENRE CANNON**

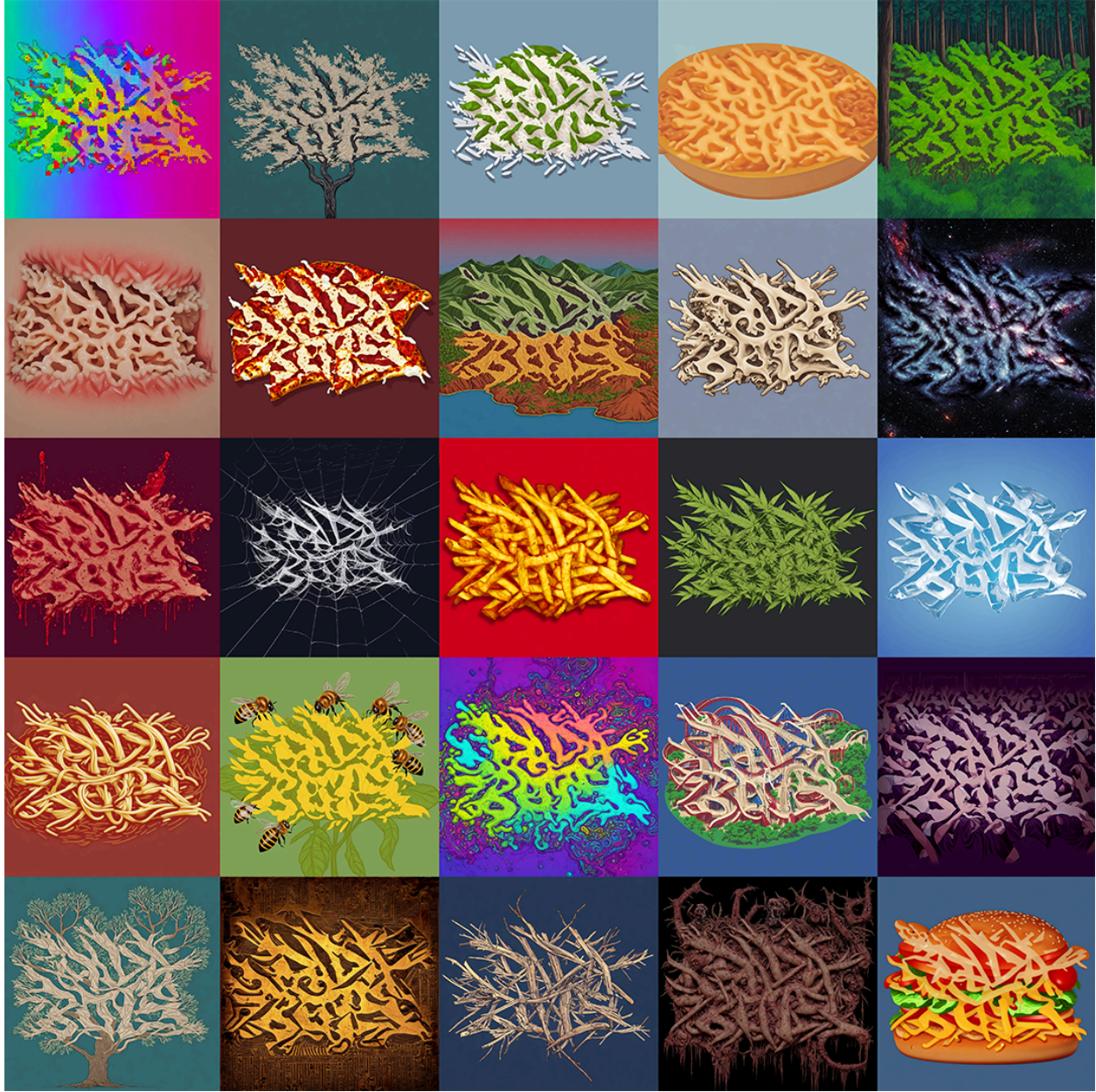
Since AI is now great at making any genre, music production is now officially all about cramming as many new genre fusion inventions as possible into one song.

This song is blatantly self-aware. Genre Cannon is a hero's journey through a multi-genre world.

## **FRENCH FRIES WERE THE INSPIRATION**

Most metal band t-shirts are black w/ an indecipherable logo. One day, CJ imagined a french fry metalhead shirt that was red, where the band logo was an indecipherable pile of french fries. Zack cheekily called it "frygoth" and we couldn't stop laughing. We made the shirt a reality using the DADABOTS logo. (All the best ideas start as inside jokes.)

It didn't stop there. Then came "weedgoth", which is our logo made of marijuana leaves. Then came others. Using Stable Diffusion + ControlNet we made 50+ variations of our DADABOTS logo as e.g. spiderweb, ice, neon, forest, etc. Each is us in a different genre. We put these logos on merch and hoodies. We put these logos on microSD cards loaded w/ music.





It made too much sense. We produce music in multiple genres. We build multi-genre machines. The logo and merch gotta be multi-genre too.

With this concept we went on a live tour across Europe, where we used our latest models to create any genre fusion the audience imagined. Every show was different. We put a phone number on the screen, and people texted in their requests. We even played at UN events in Geneva where the audience prompted us to make cultural fusions like "samba afrobeat techno".

On the last day of tour, NOXBITE reached out about wanting to do another music video. AI Song Contest was coming up soon. "YES. We should do another track together! Let's have all the logos in it, and do a million different genres" And thus the concept of GENRE CANNON formed.

In previous years of AI Song Contest, we've mixed a few genres together. The first year it was pop w/ a metal surprise. The third year it was a fusion of gospel choir, metal, and breakcore. For the fifth year we pushed it further.

## **STORY**

Our hero, dissatisfied with his asshole friends at home, goes on a journey to new worlds and their music genres. Some delightful, some terrifying, some misunderstood. Upon meeting a cthulhuian adversary in chiptune black metal world, he defeats him in a dance duel, and victoriously gains the *ability to make the genre*. Then, transcending genre altogether, he discovers how to make genre fusions, though has a mishap in his attempt (!). But in the end, he finds happiness, because of the new friends he made along the way.

## **SHIRT PORTALS**

Putting on a shirt transports him to that genre world.

The "frygoth" shirt sends him to a world of dancing ketchup and french fries. The genre is some derivation of polka. Amused by this outcome and his new friends, he tries the next shirt.

The "weedgoth" shirt sends him a world where he's having a great time getting high with dancing frogs. The genre is stoner metal frogstep.

The gore logo shirt sends him down a dark corridor where he's getting eaten alive by a terrible monster in the 1st person perspective. The genre is deathstep riddim.

The half-half shirt put him in toy world on train tracks. The toys are waving, and the genre is children's music. Then a cute style of deathcore layers in. He looks down at the bloody half of his shirt and realize the outcome won't be good. The toys wave goodbye as the train tracks send him to certain doom.



## MEGA MAN

Zack and I grew up playing the MEGA MAN games, whose concept in many ways spiritually aligns with DADABOTS. He's a robot acquiring styles, which is exactly what DADABOTS is. From our perspective, it's an allegory for machine learning. It felt obvious to make a reference. Mega Man is an android who fights with a PLASMA CANNON attached to his arm. When you learn to defeat a robot master - the boss of a level - you earn their special powers and can use their weapon. For example, Fire Man has a fire level and shoots flames. When defeated, you can now turn red and shoot flames from your cannon too. (This is where the song title comes from). But the best part of the games were the music, often stylistically matching the theme of the levels. (There are entire bands like Powerglove, Bit Brigade, and The Megs dedicated to covering the mega man OST) In some ways, MEGA MAN primed us to seek out an *algorithm of style*. We tell the story of our first neural style experiments in the documentary PIZZAFIRE.

<https://www.youtube.com/watch?v=iOy23At-bfw>

In black metal world, he puts on a Mega Man outfit. The music is dark synthwave. He's transported to black metal world, where he encounters the boss, Black Metal Man. The music switches to black metal and they fight in a duel. Black Metal Man fights by shredding w/ a BC Rich Warlock guitar. Mega Man fights back with dancing, switches to the boss's weakness (french fries), and defeats the boss. There is an epic scene where Mega Man wins the boss's powers, and it says "YOU GOT BLACK METAL" (a reference to the classic mega man games).

I wanted to reference Mega Man's melodies, but without directly ripping them off. To do this, I simply tried to write them from memory while avoiding listening to the real thing. In the end, I

was successful at making melodies that didn't plagiarize but nevertheless made me feel the nostalgia of playing the game.





Note in the earlier draft of the music video, these scenes were not rendered yet



## AI MODELS

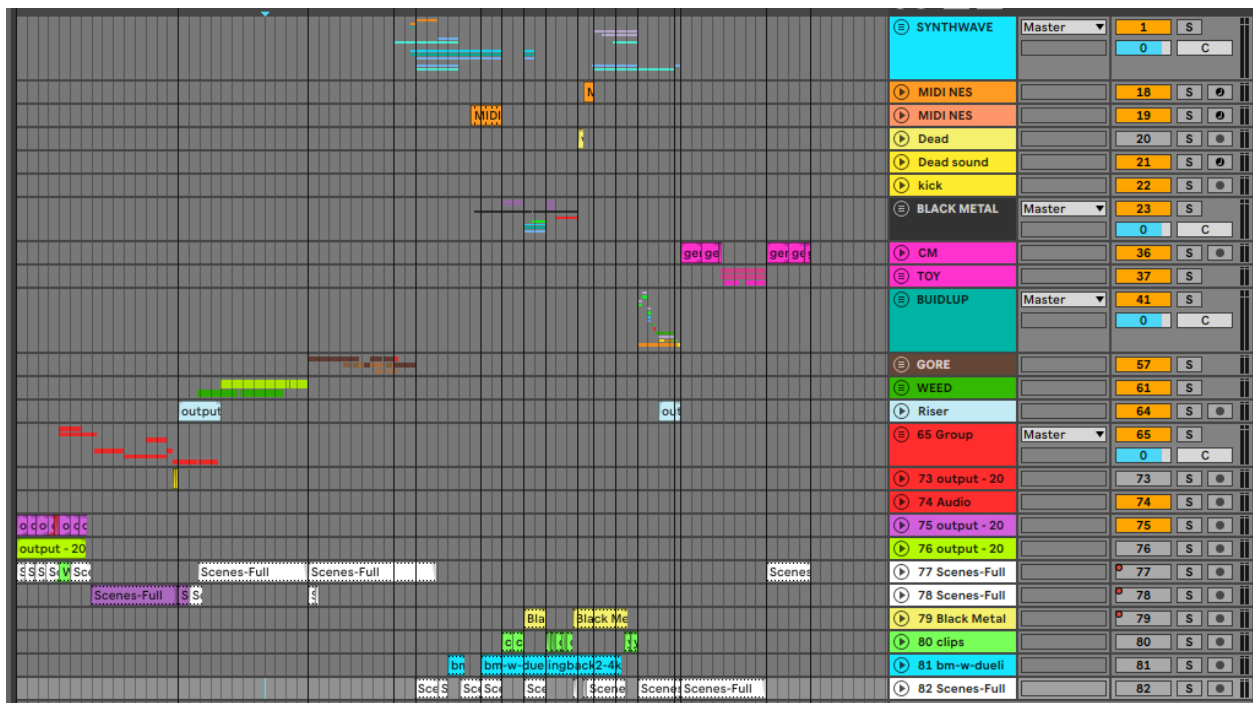
We created the Stable Audio models we used in the song (with the Harmonai team). They're audio models capable of generating sound design and music.

Stable Audio 2.0 (<https://www.stableaudio.com/user-guide/model-2>) trained on Audiosparx (<https://www.audiosparx.com/>) - It takes 10 seconds to make a 3 minute song & can do over a thousand genres & can do genre fusions

Stable Audio Open 1.0 (<https://huggingface.co/stabilityai/stable-audio-open-1.0>) trained on Freesound (<https://freesound.org/>) FMA (<https://freemusicarchive.org/>) - was released Open Source and is great at sound design and generating sound effects.

We also have alternate versions of Stable Audio that are unreleased that have other features such as CFG++

Anna Huang's Coconet (<https://magenta.tensorflow.org/coconet>) - creates harmonies from a melody and was trained on 306 Bach chorales



## PROCESS

Zack and I (DADABOTS) produced in Ableton Live w/ substantial sampling from generative models. Mahdi/Noxbite created the Music video in Blender/Houdini.

First we came up with the concept and animation storyline with Mahdi. Then we worked backwards to the music. Periodically Mahdi would send us preview clips of the video. Upon seeing the previews, we would think of music that fit EVEN better, and change it. It was for sure a dialogue between animation and music.

For example, we knew world #2 would be "weed world". And probably the music would be stoner rock or trap, two genres pretty strongly associated with weed for us. Mahdi put a ton of dancing frogs in the world. Lol. They were great. Immediately we thought "NO ITS GOTTA BE FROGSTEP".

To make frogstep we used Stable Audio to make 140 bpm beats & bass samples. We made dozens of candidates. At one point we were literally just prompting "frog dubstep". The sounds were OK. But then we tried looping some small slices of sound and it sounded exactly like frogs croaking. When Mahdi heard this he HAHHAHAHA'd in all caps.



## ARRANGING & TRANSITIONS

With multi-genre music, transitions are challenging. Genres are more than just sound, they also represent times/places/lifestyles. How do we move from one to another, naturally, without mere juxtaposition?

Overtime we developed a theory on genre modulation. We think of "genre pivots": transitory sections with things in common from both sides, which produce tension and want to resolve somewhere new. Not unlike in jazz theory when you change modes by using pivot chords.

Rhythm modulation: I think these same principles work in rhythm too. Tension rhythms (buildups, triplets, dotted syncopations) produce anticipation of resolution rhythms (four-to-the-floor, or back beat groove). Or when moving from halftime to double time (e.g. drumstep to dnb), a juxtaposition doesn't always feel natural, whereas a pivot rhythm (using dotted-eighth syncopations, between the eighth note pulse and quarter note pulse) makes the transition feel right (listen to Drumcorps' Geist it's a great example)

Timbre modulation: when the introduction of a new instrument sounds jarring, you can fix this by getting the listener to anticipate it, by gradually introducing the same frequency range right

before it. Electronic music has solved this with the riser/buildup. The riser is not only a universal signal that a change is coming, it's also preparing the ear spectrally. It wakes up the listener's high end sleeping a moment ago.

For example, at 2:25 the triangle wave & vocals are introduced before the guitar.

In 3:03 the muted snare anticipates the return of the synthwave section.

At 3:30 the "ding" anticipates the spectrum of the xylophone

Instrument pivots can also be done by switching to new instruments with similar spectral envelopes.

At 0:50 the accordion turns into an unknown instrument, spectrally halfway between a guitar and accordion, serving as transition to stoner guitar. At 0:50 you think you're still hearing an accordion perhaps more muted. But by 0:55 that note sounds more like a guitar. So at 0:58 the stoner guitar feels right layered in. And by 1:05 you don't miss the accordion anymore.

Timbre can also produce tension. For example, harsh mids or resonances. The deathstep-riddim section from 1:35-2:00 is intentionally harsh -- it's meant to make you feel like you're being ripped apart. But feels like we need a release by 2:00 and hence why we take an ambient break.

Genre modulation uses a combination of the above techniques plus more.

Sometimes the transition came to us naturally in our minds & we knew what to do. For example, from 2:10 to 2:30 the flow from synthwave to black metal was obvious: we can just drop down to a choral synth layer which fits in common.

But other times we were stumped. For example, the first minute of the song. Initially, we only had a piano loop, a polka loop, a stoner metal loop, and the frogstep. We were so stumped how to get from A to B to C to D we procrastinated on it until the very end. We had to rely on our theory of genre pivots to figure out the puzzle with the help of inpainting.

## **INTRO**

When Mahdi started us off with a clip of a man in his living room, saying "all my friends are buttholes", I knew this was going to be an adventure about finding new friends. In a way, this song is an allegory for feeling unhappy with your *home culture* & exploring what the the rest of the world has to offer.



0:00 This is a somber piano loop generated & cut up. It feels exactly like it could go straight into lofi hip hop, but we're not going there. We're going into polka.

0:14 Using the piano loop and inpainting "polka", we get an accordion melody which serves as a great transition to the polka music coming up. Our hero looks through his wardrobe for genre shirts which will transport him. He finds a frygoth shirt

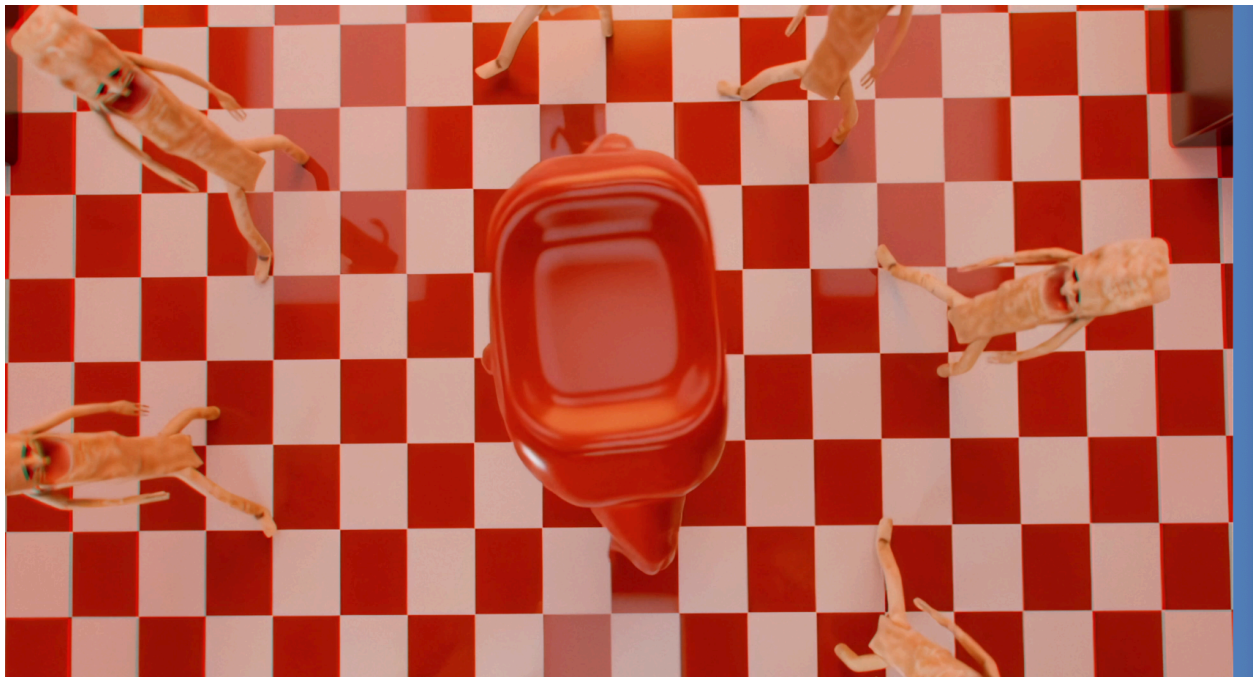
0:25 Spent hours trying to find the perfect transition right here. It's a great cadance and pickup into the energetic rhythm of the next section.

## **FRYWORLD**

0:29 Before getting to a polka drum beat, we needed to introduce the concept of rhythm without percussion

0:36 Here we prompted for polka w/ a garage-y type drum beat. Gradually introducing instruments as we build up into stoner metal.

0:48 Also spent forever searching through generated clips to find the perfect transition here.



## WEED WORLD

0:53 We had idea how we were going to get from polka to stoner metal, we saved this puzzle till last. But here we do two kinds of pivots. First the instrument pivot: the timbre sounds halfway between accordion and guitar. The first few notes seem like they could be muted accordion, but the third one seems very guitarish. Second: melody pivot. We use the melody the guitars are about to do. Thirdly we add a riser, which not only gives the expectation a drop is coming, but prepares the ear for the loudness and full frequency range incoming for the frogstep section

0:58 Because of how we set up these transitions, at this point it feels so natural to hear the stoner metal section come in here. The dirt of the power chords feels like a release of the tension. This was extremely puzzling to figure out though.

1:06 Before we go full frogstep, we introduce the timbre of the frog bass. Literally just repeating this one loop was better than the chaotic sound design we originally had. It prepares the ear to tune in. Riddim producers do this kind of thing. Also the guitars will soon be burying the frogsynth. The repetition of the same frogsynth loop seems to call better attention to it. It's how the cocktail party problem changes when a speaker starts repeating the same thing over and over.

1:14 Massive frogdrop. We've arrived at stoner metal frogstep. Usually dubstep doesn't have guitar timbres, and metal doesn't have dubstep timbres; these things kind of compete with each other, but I think the way we set up the transitions makes you accept it. The guitars, beat, and frogsynth are cut up from dozens of different generations. The riffs ended up a lot simpler than they were originally. The simplicity seems to make the fusion work better.

At 1:26 we let the frogstep stand for itself. A tension release. You can hear all the timbres that were masked by the guitar. Also we're going into another dubstep subgenre next so it makes sense to let this be itself.



## **GORE WORLD**

1:34 Gore world. This is riddim or deathstep. There's probably 20 short stable audio samples here cut up and repeated. Lots of synths, drums, guitar chugs, and growl vocals. Some of these originally came from Splice and were turned into variations using audio2audio with stable audio

1:37 Again, the repeated quarter note samples, the pauses on the 4 and 1, this is kind of a riddim production thing. I was listening to a lot of SVDDEN DEATH.

1:54 This blast beat + gated screaming combo is a total Whourkr thing that I do a lot in my music too

2:02 Ok after all that deathstep the listener needs a break. We go to a quiet section. We're about to go into synthwave so we introduce the beep layer early. This beep is a single eighth note cut from a whole loop + reverbed. To transition into the beat, we add a sheeow sound to wake up the high frequencies





## **BLACK METAL WORLD**

2:08 OK now the hero is in dark synthwave chiptune black metal world. He's dressed as Mega Man. He's gonna fight Black Metal man. The synthwave loop was generated + pitch shifted to fit the upcoming melody.

2:19 We transition from synthwave into black metal by pivoting to this dark choral synth layer, which works over both the synthwave loop and the black metal loop.

2:25 Before cutting into black metal, we do a pickup w/ triangle waves (classic mega man boss sound), black metal vocals (anticipating the genre), and a drum fill. All of it anticipates a black metal boss battle

2:28 Black metal intro riff. I wrote the melody here. I used inpainting to convert the melody into black metal guitars, and it sounds very good here.

There are generative vocals here. They are not lyrics-conditioned, so these are random syllables. Nevertheless we interpreted them in a way that fits the scene. Very surreal existential terror.

IM LOST IN MY HEAD CHANGING YOUR CONSCIOUSNESS  
SHUTDOWNNER FOR THE FIRE

DIEATH

This last word starts as DIE and then ends as if it were DEATH and I think that's funny.

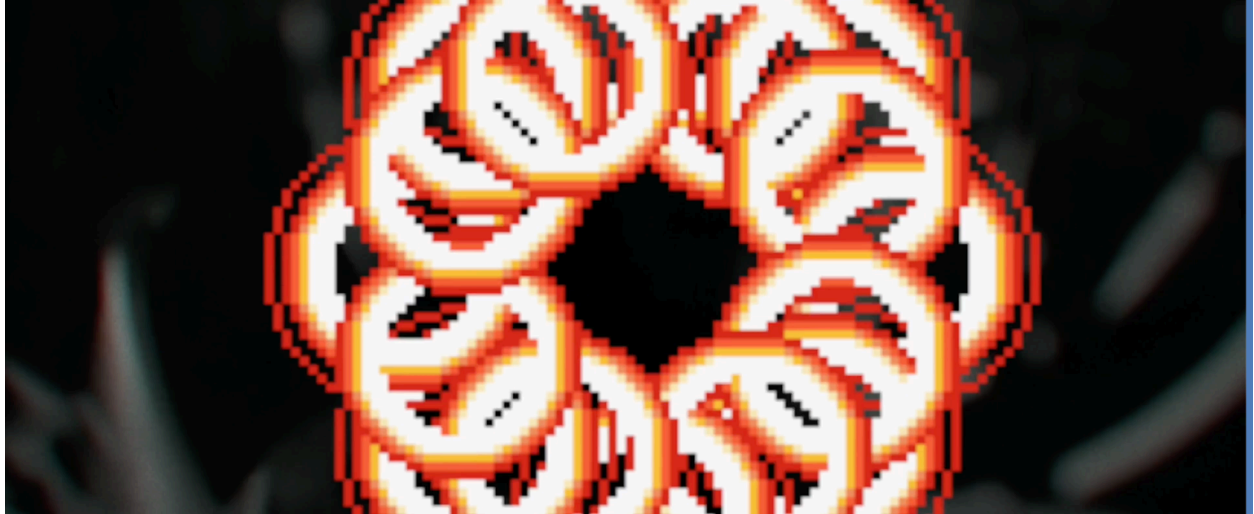


2:35 This triangle wave riff is the melody I wrote, and is what I used for inpainting and for coconet. This is mega man's dance duel theme

2:42 Guitar solo!! To make this guitar solo, I ran the melody I wrote into coconet, and generated a four part harmony. I ran this harmony into an arpeggiator, made some edits, and this created the solo. The synth is generated using Plini's Neural DSP. To call attention to the solo, we drop the beat for a couple bars and bring it back

(In the first draft of the music video, these animations were not ready yet). Mega man duels again. This time he switches powers to frygoth (accordion), which is black metal man's weakness.

2:58 The boss's health drops to zero and he is destroyed. I recreated the classic mega man "boss destroyed" sound using operator, arpeggiators, and saturation.



This melody is reminiscent of a mega man victory theme. To write a similar melody without plagiarizing, I avoided listening to any real victory themes, and tried to recall one from memory. It sounds good. Then I listened back to all the different victory themes and confirmed it wasn't exactly the same as any of them. It's kind of like I avoided overfitting by avoiding training myself too much on the melodies.

3:03 Back to synthwave. The classic snare hit pickup is so nostalgic, it's the perfect transition to signal the incoming beat. Again, here I try to write another mega man "new weapon" melody from memory. I wasn't even close to any of them, but it seems plausible. In this scene, our hero gains the ability to make black metal, much like how Mega Man gains his enemy's powers (and style) after defeat.

3:08 Here there was the harsh mixing issue that soothe2 solved. The generated pad synth loop had resonants that were masking the triangle wave melody. I didn't want to make the pads softer, nor the triangles louder. But using Soothe2 let me sculpt the resonances and save the day here.

## **TRANSCENDENT BUILDUP**

3:19 To create this section, we made a video looping through a couple dozen logos aligned to music. This was then used as the shirt texture.

The first logo is a tree, and the genre is bluegrass, playing the same melody as the previous section. The next logo is 70s psychedelic rock, playing the melody again. Then it's 90s rave psychedelic. Then tech death, frygoth, and weedgoth. It continues: rice (funk carioca), hieroglyphics (egyptian rap), sticks (ambient), crowd (trance), blood (deathstep), galaxy (jazz), and so on, rapidly speeding up, until it lands on a hybrid genre: children's music + deathcore.

To convert the melody into different genres, we initially used the init audio feature of diffusion models to noise an input & denoise it with a new prompt. For some of these genres we did this. However, to our surprise, we found that inpainting continuations made a better quality style transfer. By giving the model my input melody to start, and asking it to continue the track in a new style, sometimes it would just repeat the melody in the new style. It wouldn't always do this, but when it did, it was better quality.

Sometimes, if my input was a mixture of instruments, a continuation would perform source separation. These results got us re-imagining inpainting more as in-context learning. The input is kinda like a "context window" (like giving Claude or ChatGPT an entire book to read before prompting it).





## TOY WORLD

3:30 We started playing with a CFG++ sampler, an idea from image research, which supposedly fixes high frequency details in images (like fingers). We noticed it had given electronic music more detail, while making compositions more loose. It seems to skew denoising toward high frequency information and away from low (frequency not as in sound frequency, but latent frequency). While a typical CFG value is around 7, with a CFG++ sampler the typical value is 0.8 and you're supposed to use values between 0 and 1.

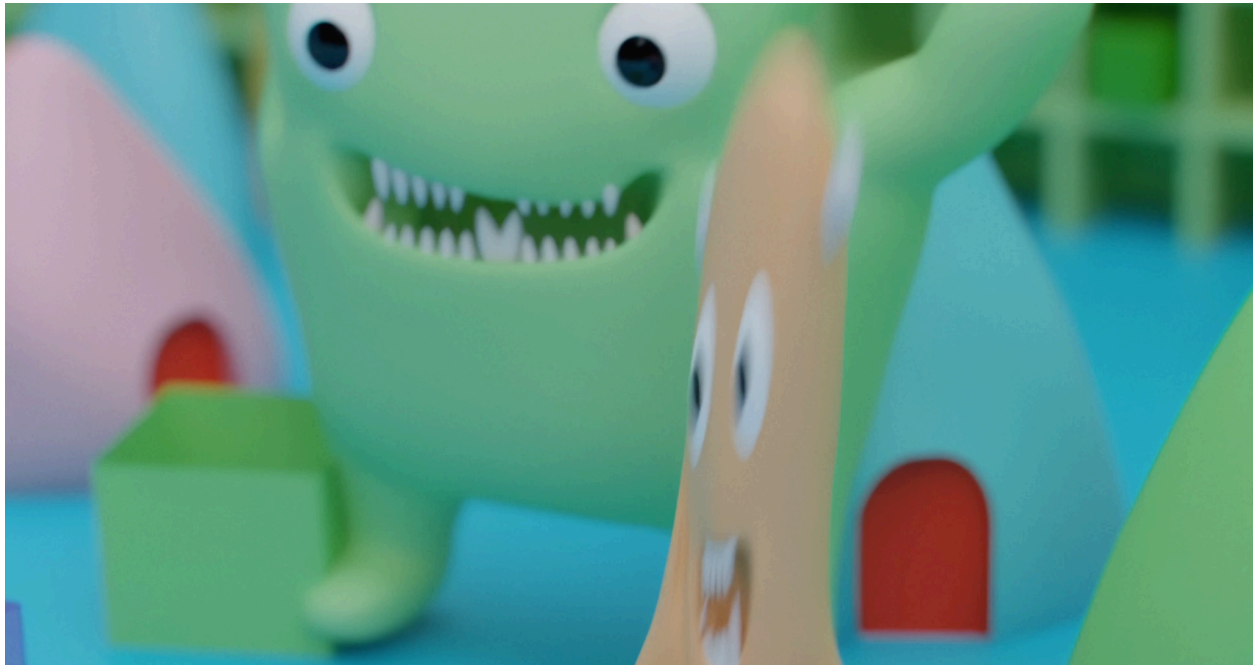
For the hell of it, I tried a value of 20, with a "deathcore" prompt, and it made this high-pitched autotune-ish screaming. I had never heard anything quite like it, like a tiny cute monster. I layered it with a +12 pitch copy and it sounded even better. I thought, "this would go perfect with children's music xylophone". I tried several "children's music" prompts until I found the perfect xylophone loop, and layered that. Ahh amazing.

Getting this result made me want to finish the song with this hybrid genre.

At first we had it at 100 BPM, but it sounded better to slow it down to 70 & remove half the bars & keep the whole song at the same tempo. It's kind of a trope to end a deathcore song with an even slower breakdown, felt right!

Again here, there are no lyrics conditioning. Instead we interpret the random syllables in a way that fits the theme—

YEAH  
I CHAIN TO EVERYTHING I CARE ABOUT  
SO GOOD  
GOOD





## **FINAL SCENE**

This may have been cut from certain versions of the video to fit under 4 minutes, but –

In the end, he finds happiness, because his worldly travels have made him wonderful new friends . . . who are, irreverently, a . . .



## MIXING & MASTERING

Finally near the end of our process, we began mixing.

Mixing generative samples is tougher than studio-recorded samples. They're generally lower fidelity. They can have random stray resonances, noise, and peaks. Extra attention is needed. Furthermore, mixing multi-genre music is challenging. Every genre has its own way of sounding, its own expected frequency profile, loudness, and dynamic range. For example, modern electronic music should utilize the full range of a sound system, with crisply defined instruments. Whereas black metal should sound subdued and buried. Yet all these genres need to fit together into one experience. How do you blend them together cohesively?

The listener shouldn't feel they need to change the volume within the track. So a certain amount of evenness across the track is desired. But you also need range. Certain genres (like deathstep) depend on a LOUD drop in contrast to its surrounding parts. The french polka shouldn't be the same loudness as the deathstep drop. Balancing evenness and range here is quite the puzzle.

We used the standard mixing toolkit: multiband compressors, limiters, harmonic saturation, EQ, stereo widening, reverb. To separate instruments you give them each a home in the space of frequencies & stereo-width (in my mind I see sound as a 3D space I'm flying through). Giving

sounds to different reverbs makes them feel like they're in different environments, which is another way to separate them. If an instrument is too buried, high frequency excitation can make an instrument more present.

Metal guitars have noisy mid-high spectrums, and don't always blend well with the pure tones of synths. Adding saturation/reverb to synths (or on a bus combined with the guitars) gives them more noisy color, which blends them together better. The visual analogy I think of is you're applying an instagram photo filter to them, so they feel together in the same scene.

Unruly transient peaks are tamed with compression and limiters. Generative music loops seem already pretty decent with their dynamics and don't really need too much compression. But sometimes multi-band compression is there as a mixing choice. For example in the black metal section of 2:30, because black metal is a buried aesthetic, adding full band compression made the drums nearly disappear. Applying compression to just a single band made the guitars stronger without affecting the drums.

A secret weapon I use is soothe2. This plugin is awesome. It lets you shape resonances. Often it's the case that two instruments sound good alone, they're both in key/rhythm, but when you mix them together they clash timbrely. Parts of one may hide the other (frequency masking). Sometimes this is fixable with EQ. But what if they're in the same EQ range? Sometimes this is because of stray resonances. Soothe2 squashes them. A good example is at 3:10 when the synth chords come in. Those chords were interfering with the triangle wave melody. But I didn't want the chords softer, nor the triangle waves louder. Soothe2 saved the day.

Mastering is also not straightforward with multi-genre music, because what should you use as a reference track? Should the electronic music parts be as loud as the rest of electronic music? I lean toward no. Moreso I'm thinking an album (or set) of multigenre tracks should just be tweaked to fit with each other.