SINGER/SONGWRITERS WITH JUKEBOX

- I've been playing around with jukebox for about a year and these are some tips and tricks I've picked up along the way. This is not a definitive guide! I think of Jukebox as an instrument I am learning to play, no different from piano or guitar. Sure, some Jukebox outputs sound more musically pleasing than others, but as long as the program is up-and-running – and generating music – there is no real "right way" thereafter.
- I write my lyrics, then have Jukebox "perform" them. After that I do the post-processing myself on Audacity… the entire chain is open source.
    - My YouTube: https://www.youtube.com/channel/UCVRpMo19NwYKloFhnw6QzMg
    - My SoundCloud: https://soundcloud.com/songshtr/sets/machinecroon
- If you're starting off, the best place to begin is the OpenAI paper itself, and the sample explorer to see what is "achievable". Also worth skimming through the "multi artists/genres" portion of the paper, which I implemented in my notebook and is central to my process.
    - https://openai.com/blog/jukebox/
- The next place to go is the discord channel run (primarily) by Brocaloo. This contains the latest Colab notebooks that you will need to run Jukebox. The community is extremely vibrant and no question will remain unanswered for long!
    - Discord Channel: https://discord.com/channels/766622617393430559/769885769891840010
- The discord channel contains a few different notebooks that you can use. Most of them run on the free version of Colab, but I signed on for Colab Pro (~$10/mth) to get access to more RAM and faster GPUs. Get started by watching a couple of Youtube tutorials on how these notebooks work (I won't go into those details here). Once again, the folks on the discord channel are extremely helpful, and welcoming to starters with questions:
    - Tutorial 1 (Brocaloo 1 click notebook): https://www.youtube.com/watch?v=CCmTAFe5GcU
    - Tutorial 2: https://www.youtube.com/watch?v=Z1E5cu5sPo0
    - Tutorial 3 (Johannez… I based my notebook upon his): https://www.youtube.com/watch?v=5wn3htQI4JA
- The main change I made from the base notebooks was to implement the "multi artists/genres"… i.e. create music in a weighted style of 2 artists/genres, with different weights. This was discussed in the original paper, but the code was never released. However OpenAI discussed how to implement this on Github and I was able to get a version of the notebook going. I wanted the machine to SING… ie be lyrical, vocal heavy, not electronica… but wanted to see if I could get voices to blend. This "works" but not fully – Johnny Cash, just like in real life, will always make any song his own.
    - Notebook: https://github.com/songeater/jukebox
- The Jukebox model is trained on raw audio. Roughly speaking, during training, raw audio is "compressed" in 2 steps, then the compressed tokens are fed to the learning model (a transformer) in the 3rd step. During generation, the process is reversed. You feed in a primer song-snippet, specify artists/genres, add lyrics, and Jukebox generates a VERY rough version of the song (called "Level 2"). This rough version is then upscaled twice (i.e. the reverse of the 2 compression steps, first into "Level 1" then into "Level 0"). This brings me to my first real tip:

  - During generation, you have to upscale once to Level 1 (or else you won't have anything "listenable") but it's just fine to skip the second upscaling step to Level 0. The second upscaling step is – by far – the most time-consuming step, and yes your sample sounds somewhat better after it – but the relative difference between Level 0 and 1 is nowhere near the difference between 1 and 2. The sound is still lo-fi / scratchy and a lot of it can be "fixed" in post-processing using simple low-pass filters).
  - For example, to generate ~4 samples on a P100 takes about 2 hours on the generation step and another 2 hours on the first upsampling step. But the second upsampling step would take another 8 hours!
- "Embrace the scratch." Jukebox generates a mono WAV file with instruments/vocals all smooshed into a single-track. While the model is quite powerful, I liken it to a great band that showed up at a studio session after a couple of hours at the bar. The band can sing and shred with the best of them, but the band often gets bored, distracted, loose. Also the recording equipment in the studio was last upgraded in the 1950s! It is what it is… embrace the mess and the lo-fi sound. Given what I was looking to do, I was just fine going with the "tapes-found-in-the-attic" vibe… and this definitely allowed me to feel just fine about abandoning the last upsampling step. YMMV if you are looking to generate electronica for example.
- Picking primers: For generating vocal sequences, I've found that primers are best if they are a) relatively short (3-5 seconds) and b) contains a relatively clear beat or chord sequence. A full orchestral primer leads the model to generate too many instruments from the get go and tends to muddy up the vocals.
- Picking artists: There is a joke on the Discord channel that "Ed Sheeran will sing anything." While Ed was not quite what I was going for, finding an artist with a large catalogue with lots of vocals will typically generate lyrics well… eg Johnny Cash, Frank Sinatra, Leonard Cohen… check out the Jukebox sample explorer. I typically blend two artists and leave the genre as "unknown" because I don't want to constrain the model too much (specifying 2 artists and 2 genres tended to lead to more "garbling").
- After that… it's a volume game! You have to generate a lot of samples, then splice and paste together good sections to form a song. The most important thing is to keep the primer consistent … as long as the primer is consistent, Jukebox will create samples with the same beat and key, allowing different samples to be welded together. The artists/genres/lyrics can vary… but the primer is the glue that binds them together.
- Jukebox generates between 3-4 samples per run/batch (it can do more, but requires more memory than Colab Pro will give you). Each batch takes about 4-6 hours based on what GPU you get… so I typically don't do more than 1-2 batches a day. A ~2-3 minute song can take anywhere between 15-100 samples to generate. So yes… it can be a week before you a rough cut of a song down.
  - Sometimes you get lucky and the entire song comes out in one go! https://www.youtube.com/watch?v=huq37LWvURc
- Keep a spreadsheet! It really helps to keep a spreadsheet that tracks the primer/artist/genre combos and also a quick "rating" of each sample generated by the combo. I tend to save most of the raw WAVs too… unless the batch was completely unusable. Tracking really helps you later as the volume of your raw sample grows and you start to forget what "worked."

- Once you get a bunch of samples… it's a lot of listening / cutting / pasting / moving around in Jukebox until you have a mono-track of a sing you like. After that you get to post-processing and… well, there is no end to it. I am FAR from any sort of expert on mastering… but again, here are some things that worked for me:
    - Jukebox tracks are hissy to clean them up by applying high-pass filters <100hz and low-pass filters >10khz.
    - I typically duplicate a mono track after cleaning up, then use heavy compression on track and a small delay. Then push the original mono track to the left channel, the compressed channel to the right channel… and voila, you have a VERY poor man's stereo track.
    - One I have the stereo track, feed to Spleeter (I use the desktop GUI) to separate the vocals from drums, bass, and other. https://makenweb.com/SpleeterGUI
    - Definitely clean up the "vocals" buy applying more aggressive high/low filters, then tinker along as far as you can go. Lots of EQ and compression… "to taste."
    - I don't have monitor speakers at home, so I mostly master using headphones but also try to listen to songs on very different systems: eg TV, tinny computer speaker, iPhone+AirPods
- And then it's done… mix the track and release it out to the world! The first time I did this, I felt like I had discovered a superpower.