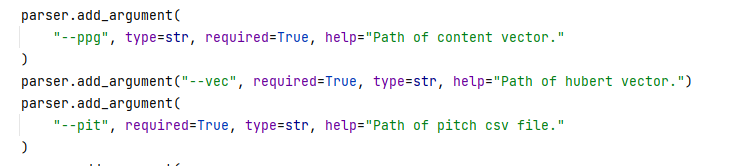
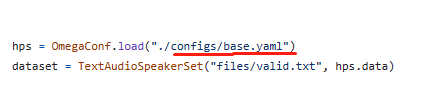
1 & 2.

Question: What's the difference between this content vector and this Hubert vector? I see in your application that "required = True" is stated, does your library include a feature extraction process? And are there any requirements for slice sizes? What is the function of preprocess\_zzz under preprocessing, and there seems to be an absence of base.yaml.





Answer: The preprocessing can be directly done using whisper-sovits svc. It's identical.

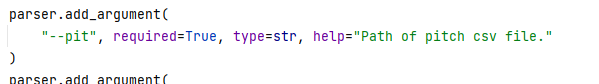
3.

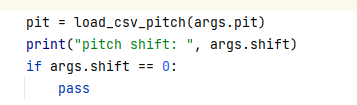
Question: Why did only directories get created but no features were stored after executing the preprocess.py file?

Answer: Ah, when I have time later, I will upload a version that can be run directly. It seems like there have been some modifications to file names and such during my own experimentation.

4.

Question: Does this need to be converted into csv format next? What I've generated using whisper is npy. What format do you guys load the pitch data in? Right now it seems to be loading in csv.



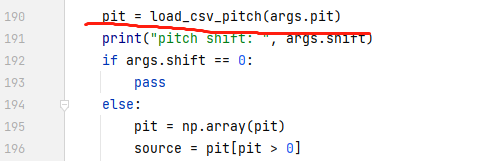


Answer: This might be something the author of whisper-sovits svc forgot to change? I hadn't noticed. The pitch data loaded is also in npy format. You can check under 'dataset' where this is mentioned.

5.

Question: In the version on your Github, svc\_inference.py writes --pit as csv while its main() function works differently. Consequently, it returns an error message telling me "UnicodeDecodeError: 'utf-8' codec can't decode byte 0x93 in position 0: invalid start byte": could you help me figure it out whether I passed the parameters wrongly or perhaps the version on Github differs from what you use?

I'm currently using -- ppg --vec --pit with path directed towards the npy files.



Answer:

The training phase uses npy for pitch while testing uses csv. You can opt to pass a new path as the inference script would generate one accordingly. Referring to my script should help, which keeps test data in a separate directory.



6.

Question:

What's the dimensionality of spk\_emb for the unfrozen 1700 version?

Answer:

256

7.

Question:

Do you have any more convincing audio examples generated using Samoye? Ideally ones with both original and synthesized voices for comparison? Can you let me listen?

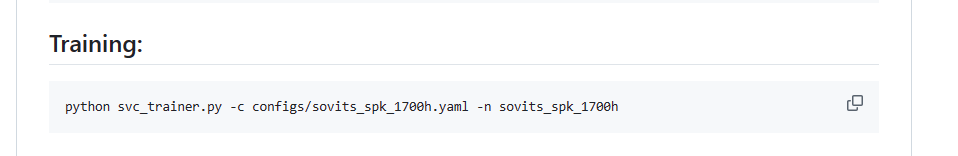
Answer:

The GitHub homepage contains audios of animal sounds transformed into human singing, along with one male one female transformed audio samples. The first part is taken from reference audio (target) and then followed by transformed audios. This setup is similar to that involving cats and dogs: original audio first followed by synthesized.

8.

Question:

While attempting this segment today, I encountered a missing file error stating "FileNotFoundError: [Errno 2] No such file or directory: 'files/train\_filtered\_1700\_withoutnan.txt'". May I know what's this file about and am I supposed to write it myself?



Answer:

Ah... You should obtain a train.txt once preprocessing concludes; just use that directly!

9.

Question:

[Embarrassed] Primarily because I haven't used whisper before so I'm not very familiar with it. At times when looking in comparison, I can't tell whether certain parts have been modified by you guys.

Answer:

Quite a lot of modifications were indeed made [Blushing] Plus a portion of the code that has been uploaded was actually trimmed off.