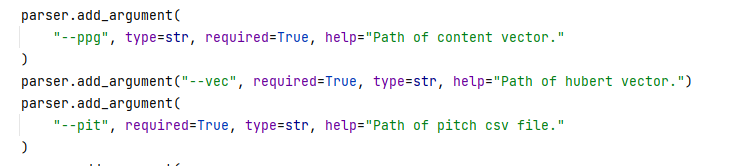
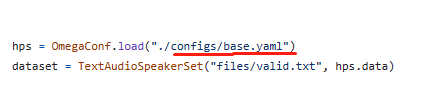
1&2、这个content vector和这个hubert vector有什么区别啊



因为我看你这个上面写的是 required = True ，你的这个库里面包含提取特征的部分吗？对切片大小有没有要求啥的

、这个preprocessing里面的preprocess\_zzz是做什么的文件啊

没有这个base.ymal



答: 预处理那里可以直接下载whisper-sovits svc弄 一样的

1. 还有一个问题就是为啥我跑完preprocess这个py文件之后光创建了文件夹 什么特征都没存啊

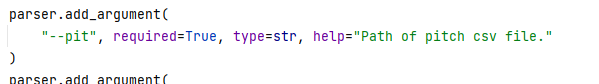
答: emm，我后续有空直接把能直接跑的版本放上去吧，其实都是我做实验的时候改了些文件名之类的

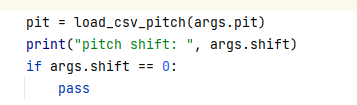
4、问题:

这个要再转成csv吗？我用whisper那个生成的是npy

那你们load的pitch data是什么格式呀

现在他load的应该是csv





回答:

这个可能是whisper-sovits svc作者忘改了？我没注意过这个哦

load的pitch data也是npy

这个可以看dataset里有写

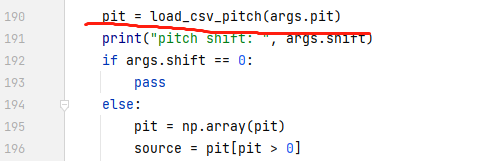
5、

问题:

你们github上那个版本 svc\_inference.py里面的--pit 写的是csv然后这个文件里的main()函数是这样写的

然后我就会报错 提示我 UnicodeDecodeError: 'utf-8' codec can't decode byte 0x93 in position 0: invalid start byte

你可以帮我看一下不？ 是我传参传错了还是现在github上的版本和你们用的不一样啊？



我现在用的-- ppg --vec --pit 传的路径是npy的路径

回答:

这个train的pitch是用的npy

然后test用的是csv

你可以直接传个新的路径，inference里面会用脚本生成

可以参考我这个脚本, 单独给个文件夹放test的数据



5、

问题:

还有一个问题是你这个库里这个infer的部分

python svc\_inference.py --config configs/sovits\_spk\_1700h.yaml --model sovits\_spk\_1700h\_0020.pt --spk spk.wav --wave content.wav

spk输入的是客人音色的音频， --wave 是原唱干声的音频对吧？可是我迁移出来时候 音频内容也跟我目标用户音色的一样，而且我用这个svc\_inference的时候 还要传--ppg , --vec --pit --shift 是这个ppg vec和pit要传入原唱的吗？我有点没搞清楚

回答:

nfer的代码里面, --ppg , --vec --pit --shift 是这个ppg vec和pit. 这些参数哪些传入原唱的 哪些传入目标音色的

限制max是30秒的音频, 是否可以改, 改之后权重是否不能用

这个解答后稍后更新到GitHub

6、

问题:

这个1700解冻版本的spk\_emb的维度是多少呀？

回答:

256

7、

问题:

有没有用samoye得到的效果比较好的音频啊？最好有原本声音和生成的声音。可以给我听一下不？

我看到了首页上的那两个音频 有没有提取音色用的音频呀 我想对比听一下

回答:

github首页上放动物作为reference audio转人类的歌声, 也放了一男一女作为reference audio 转人类的歌声.

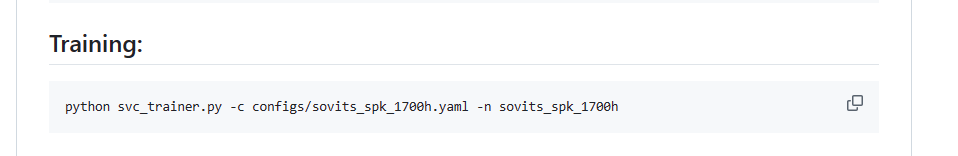
先放的提取音色用的音频（目标）再放的转换后音频。拼在一起了, 这5个音频.

和猫狗那个一样，前面是原音频, 后面是生成的

8、

问题:

我今天尝试这个部分的时候报了缺少文件的错误



FileNotFoundError: [Errno 2] No such file or directory: 'files/train\_filtered\_1700\_withoutnan.txt'

请问这个文件是干什么的 需要我自己写一个吗？

回答:

额，那个preprocess弄完有一个train.txt

你直接用那个就行了

9、

问题:

[尴尬]主要是我之前没用过那个whisper的那个 还不太熟悉

有的时候对照看 有的时候不确定是不是你们改动过的地方

回答:

确实改动挺多的[捂脸]而且放上来的有一部分代码删减掉了