1. 客观指标对比（WER, CER, STOI, MCD, F0 RMSE）
   1. 主要需求是写出各个指标的计算方法，输入分别是**模型生成的mel或音频**和**GT mel或音频**，(可以参考其他语音合成论文的开源代码) 例如：

def get\_f0\_rmse(f0\_pred, f0\_gt):

return np.sqrt(np.mean((f0\_pred - f0\_gt) \*\* 2))

…

1. 数据集制作 (民歌/美声数据集，M4Singer)
   1. 民歌/美声数据集
      1. **整理数据路径**，最好按照下面的方式安排文件

民歌美声完整音频/民歌/性别/xx歌手(如果没有名字的话就用性别+数字代替，例如：男-赵四，男-1，女-翠花，女-1，…)/歌曲名（不包含人名）.wav

* + 1. 对**每首歌进行切片**，可以使用whisper3等语音识别模型获取每首歌的歌词，和对应的时间戳，最好每个切片控制在4-10秒之间（可以丢弃低于2秒的数据），可以保存对应切片的歌词为同名的.txt文件，切片命名如下：

示例：民歌美声数据集/audio/美声/男-赵四/歌唱祖国/001.wav

（民歌）

民歌美声数据集/audio/民歌/xx歌手/xx歌曲/001.wav

民歌美声数据集/audio/民歌/xx歌手/xx歌曲/001.txt

民歌美声数据集/audio/民歌/xx歌手/xx歌曲/002.wav

民歌美声数据集/audio/民歌/xx歌手/xx歌曲/002.txt

…

民歌美声数据集/audio/民歌/xx歌手/xx歌曲/0XX.wav

民歌美声数据集/audio/民歌/xx歌手/xx歌曲/0XX.txt

（美声）

…

* + 1. **粗略标签+大模型Prompt，**可以仿照Libritts-P的写法，给出每个切片的一些数据，例如

Style Prompt Key：民歌-男/

Style Prompt：

* 1. M4Singer数据集
     1. 还没想好，大概就是仿照libritts-p的写法，加上一些歌手技巧的注释，或根据NISQA模型得到预测的MOS得分值，再通过大模型模板得到自然语言描述