

gst帮助tacotron韵律建模

目标：

tacotron在直接使用text合成audio的时候，可能存在韵律不佳的情况，例如重音念错，断句不自然等，而使用了gst帮助建模后可以改善问题。

假设：

1. 利用 (text , style-embedding) 的训练数据，从text学习得到gst信息
2. 将学到的gst信息和原来的text-embedding按照同样办法拼接，输入gst-tacotron的decoder继续生成audio，其prosody会更好

问题验证：

1. tacotron在中文（或者英文）数据集上的模型，合成的audio有韵律不正确的情况
2. 通过gst帮助建模的tacotron比前者问题有改进

计划

1. 完成对中文数据的分类工作
 - a. 比较拼音的声调拆分前后作character 的实验效果
2. 验证tacotron在测试集的不好的情况
 - a. 中文-xijunm?
 - b. 英文-ljspeech blizzard等训练数据
3. 思考设计gst的方法
- 4.