Lab 11: Speech synthesis

1 Intro

本次实验旨在使用训练好的语音合成模型合成语音，体会语音合成流程和效果

2 FastSpeech 2

虽然FastSpeech作为一个non-autogressive TTS模型已经取得了比auto-regressive模型如Tacotron更快的生成速度和类似的语音质量，但是FastSpeech仍然存在一些缺点，比如（1）使用一个auto-regressive的TTS模型作为teacher训练模型非常耗费时间；（2）使用知识蒸馏的方式来训练模型会导致信息损失，从而对合成出的语音的音质造成影响。
在FastSpeech 2中，作者针对这些问题进行了改进，作者首先摒弃了知识蒸馏的teacher-student训练，采用了直接在ground-truth上训练的方式。其次在模型中引入了更多的可以控制语音的输入，其中既包括我们在FastSpeech中提到的phoneme duration，也包括energy、pitch等新的量。作者将这个模型命名为FastSpeech2。作者在此基础之上提出了FastSpeech2s，这个模型可以直接从text生成语音而不是mel-spectrogram。实验结果证明FastSpeech2的训练速度比FastSpeech加快了3倍，FastSpeech2s有比其它模型更快的合成速度。在音质方面，FastSpeech2和2s都超过了之前auto-regressive模型。

3 TODO

- 使用提供的FastSpeech code (`https://github.com/ming024/FastSpeech2`)及训练好的模型(`https://drive.google.com/drive/folders/1DOhZGlTLMbbAAFZmZGDdc77kz1PloS7F?usp=sharing`)，合成语音（至少5条，中文、英文均可）
- 上交实验报告，包含：详细实验步骤，FastSpeech 2 模型理解。并打包合成的音频一起上交。
- 附加题：可结合Lab 10 及课上所学内容，画出频谱图。

Submit

- 2021xxxxxx_xiaoming_lab11.zip (./audio_samples ./report.pdf)
- Email yihanwu@ruc.edu.cn, DDL 2022.12.25 20:00