**Lifelong Learning + Tacotron2实验**

1. 目标：不想传统MultiSpeaker-TTS那样用多个Speaker同时进行训练，这个模型一次只有一个Speaker进行训练，且随着Speaker的不断增加，模型不会忘记之前Speaker的发音；
2. Tacotron2 Encoder部分的Embedding和Speaker的one hot编码concentrate，选取one hot编码的维度为128；
3. 首先使用LJSpeech数据集进行训练，将Speaker的one hot编码为[1, 0, 0, …, 0]；
4. 当合成效果比较好之后，从VCTK选取另外一个Speaker进行训练，这时这个Speaker的编码为[0, 1, 0, …, 0]，以此类推；
5. 使用方法Elastic Weight Consolidation（EWC，<https://arxiv.org/abs/1612.00796>）进行Lifelong Learning，以保证模型不忘记之前Speaker的音色；
6. 以此种方法训练5个人，验证Lifelong Learning的可行性。

改进：

1. 如果一个人一个人这样子来训练的话，会导致模型无法区分Speaker Embedding和Character Embedding的区别，所以在初始阶段还是需要有多个人就行训练，准备在VCTK数据集中选取50个人就行训练，再用剩余的数据one by one地进行训练；
2. 直接用one hot编码进行concatenate训练的话，Speaker Embedding的信息不足，数值上也入character embedding差距较大，准备还是使用一个learnable的embedding进行训练。