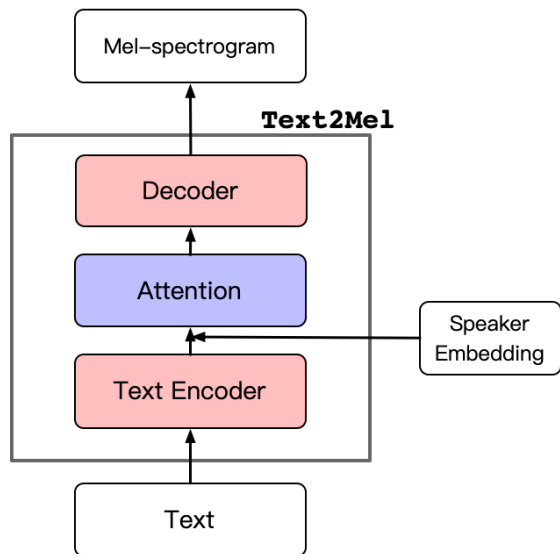


Hieratron: Hierarchical Speech Synthesis Model

声音克隆流程

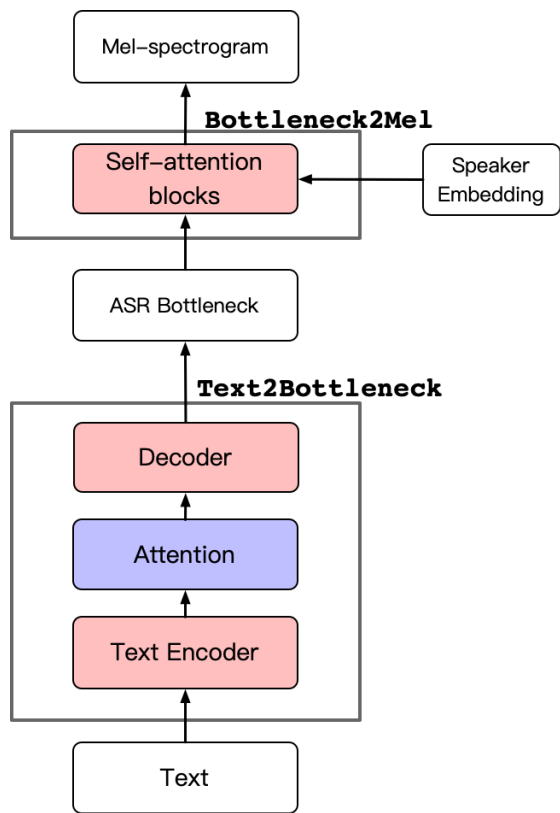
- 先使用多说话人数据预训练平均模型，再使用目标说话人Fine-tune平均模型。
- 问题：
 - 极少量数据下进行声音克隆（训练平均模型的数据越多，需要的目标说话人数据越少）
 - 克隆音色的同时加上其他控制



改进思路

使用分层的模型，模型的不同层建模声音的不同特征。先建模内容、时长、风格等信息；再建模音色信息。实现音色与其他特征硬分离。

模型结构



说明

- 实现音色与其他信息相分离，实现Cross X Voice cloning.
- 使用不同的数据训练模型的不同部分，在极少量数据下实现稳定的声音克隆。
 - bottleneck2mel 不需要任何标注信息，很容易收集海量说话人数据