第26组 第四周（4月22日-4月28日） 大作业周报

1. 本周大作业完成情况

队内成员，通过运行并比较相关领域的各种模型，以及各种相关音频增强的模型后，发现涉及音频的模型的训练周期和时常感人，模型大小十分可观。

同时团队成员十分重视的一点是文本预处理当中的对齐操作，模型是如果是隐式对齐，也就是模型自动学习到的，这一步跟课程关系很紧密很重要。

**显式对齐：**文本和音频之间的对齐是显式进行的，这意味着需要额外的步骤或数据来确保文本中的每个音素或字词与音频信号中的相应部分相匹配。

**隐式对齐：**在更现代的端到端TTS系统中，如LightSpeech和FastSpeech系列模型，对齐通常是隐式完成的。这些模型被设计为直接将文本输入转换为语音输出，无需显式的对齐步骤。它们通过学习文本表示和声学特征之间的复杂关系来实现这一点。在训练过程中，模型会自动学习如何将文本序列映射到相应的语音波形，从而隐式地处理对产问题。

同时考虑到显示对齐的方式训练模型，对数据质量要求很高，因此初步考虑采用轻量级的TTS模型。

1. 成员具体分工情况

|  |  |  |
| --- | --- | --- |
| 姓 名 | 学 号 | 主要工作 |
| 唐友朋 | 2114050025 | * 阅读文献：   1. [Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers | Papers With Code](https://arxiv.org/pdf/2301.02111v1)  2. [LightSpeech: Lightweight and Fast Text to Speech with Neural Architecture Search](https://arxiv.org/pdf/2102.04040v1)  3. [NaturalSpeech: End-to-End Text to Speech Synthesis with Human-Level Quality](https://arxiv.org/pdf/2205.04421v2) |
| 孟天宇 | 2116020016 | * 尝试构建TTS训练数据集 * 系统性地学习并掌握了seq2seq模型的基础框架 * 学习运行并使用bark预训练模型实现文本to语音 * 阅读文献：   1. [Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers | Papers With Code](https://arxiv.org/pdf/2301.02111v1)  2. [LightSpeech: Lightweight and Fast Text to Speech with Neural Architecture Search](https://arxiv.org/pdf/2102.04040v1)  3. [NaturalSpeech: End-to-End Text to Speech Synthesis with Human-Level Quality](https://arxiv.org/pdf/2205.04421v2) |
| 张歆悦 | 2113060072 | * 阅读文献：   1. [Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers | Papers With Code](https://arxiv.org/pdf/2301.02111v1)  2. [LightSpeech: Lightweight and Fast Text to Speech with Neural Architecture Search](https://arxiv.org/pdf/2102.04040v1)  3. [NaturalSpeech: End-to-End Text to Speech Synthesis with Human-Level Quality](https://arxiv.org/pdf/2205.04421v2) |

1. 遇到问题及解决方案

团队成员仍有基础音频教程未完成成学习，下周互相督促尽快完成基础知识地学习，避免后期论文及代码相关工作掉队。

1. 下周工作计划
   1. 针对本周阅读的论文选择比较后，开始着手复现工作
   2. 孟天宇开始代码复现工作，唐友朋和张歆悦继续阅读论文，提出问题，孟天宇跑通后给提供requirements文件，然后一起看代码，不懂互相问。
   3. 争取12号前将基本地TTS功能实现后，争取尝试互信息