人工智能与数据挖掘课程设计

报告题目

**组员信息：**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | 姓名 | 学号 | 成员贡献 | 实践成绩 |
| 1 | 黄景亨 | 2022212378 |  |  |
| 2 | 李兆耀 |  |  |  |
| 3 |  |  |  |  |
| 4 |  |  |  |  |

# **引言**

近年来，人工智能在语音合成与声音克隆领域取得了突破性进展。作为一项充满潜力的技术，声音克隆与语音合成不仅应用于语音助手、虚拟主播、个性化语音导航等领域，还在文化传承（如方言保护）、教育和娱乐等方面展现了广泛的应用前景。为了探索和实践这一技术，我们计划设计并实现一个具备声音克隆与文本转语音能力的项目。

**关键功能:**

①输入声音:使用深度学习算法构建声学模型。

②输出声音:使用构建好的模型,将文本或旋律转换为目标语音输出。

**实现过程:**

一、声音的获取与预处理。

二、克隆声音:

使用VITS 框架和神经网络算法构建和训练目标声音模型。

三、文本转语音:

使用训练好的模型输出指定的文本朗读音频。

**项目目标:**

实现已有的AI音色克隆软件“GPT-SOVITS”的基本功能(使用素材生成音色并且能文本转语音)并在此基础上增加方言转换功能。

# **实验数据**

## 数据获取或预处理

通过pr等剪辑软件分离出目标原声素材,再用vocalremover等网站消除背景音，保留人声,最终获取一份能反映目标声音特征的音频文件。

## 数据统计

**①标注**：包括音频的分段（分离清晰语音和噪声片段）、文本标注（对应发音文本）和特征提取（如F0频率、节奏）。

**②特征提取**：

音频特征：使用 Mel 频谱等表征音频特性。

声学特征：提取韵律、音高、时长等信息。

# **方法/模型**

主要参考SoVITS大模型:

重建音频

基本原理:低相似度音频---------------------------------->高相似度音频

↑

|

|包含音色的 token

|

目标音频

# **结果评测/展示**

方法评测或功能展示。

# **总结**

工作总结、局限性分析。