可控语音合成系统基于FastSpeech2模型的任务需求书

项目背景

目前，自然语言处理和音频处理领域内有许多高质量的自动语音合成（TTS）系统，但很少有系统同时提供音色、情感和音调的可控性。FastSpeech2模型以其高效性和高质量的生成能力而获得认可。本项目旨在基于FastSpeech2模型进行改造，实现一个音色、情感、音调可控的语音合成系统，并通过图形用户界面（GUI）简化用户操作。

目标

1. 音色可控：用户应能自由选择或调整音色。
2. 情感可控：用户应能注入特定的情感，如高兴、悲伤或愤怒。
3. 音调可控：用户应能调整语句的音调。
4. GUI操作：提供直观易用的图形界面，使非专业用户也能轻松使用。

功能需求

1. 模型训练与微调：对FastSpeech2模型进行改造和微调。
2. 用户界面

技术路线

1. 数据准备：AISHELL3 数据集或 baker
2. 模型改造与训练：在FastSpeech2的基础上添加可控制的模块，使用标记的数据集进行模型训练。
3. GUI设计与实现：Django

技术依赖

1. 深度学习框架：Pytorch2.0.1
2. GUI框架：Django + Bootstrap