南华大学

计算机学院

（ 2022秋季学期 ）

《软件综合实训II》报 告

|  |  |  |
| --- | --- | --- |
| 小组成员  姓名 | 学号 | 班级 |
|  |  | 20软工 班 |
|  |  | 20软卓 班 |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |
| --- | --- |
| 指导教师 |  |

**目录（竞赛类根据此目录整理内容）**

[一、 小组成员信息及分工 3](#_Toc23856)

[二、 选题分析部分 3](#_Toc23718)

[1、选题一：A01智能云电子警察系统 3](#_Toc19779)

[1.1题目概述 3](#_Toc21860)

[1.2所需技术和数据 3](#_Toc31301)

[1.3 国内外有关研究现状 4](#_Toc12869)

[1.4 技术/算法分析 5](#_Toc15121)

[1.5 可行性分析与技术难点 8](#_Toc24633)

[2、选题二：A05视频人像分割算法 9](#_Toc31817)

[2.1题目概述 9](#_Toc26350)

[2.2要求及所需技术 9](#_Toc1589)

[2.3国内外有关研究现状 9](#_Toc30635)

[2.4技术/算法分析 10](#_Toc11923)

[2.5可行性分析与技术难点 14](#_Toc15469)

[3、选题三：A14赋能垃圾分类真正落地见实效的信息系统 15](#_Toc15837)

[3.1 题目概述 15](#_Toc12421)

[3.2 所需技术和数据 15](#_Toc19411)

[3.3 国内外有关研究现状 15](#_Toc22405)

[3.4 技术/算法分析 16](#_Toc13646)

[3.5 可行性分析与技术难点 21](#_Toc3140)

[4、选题四：A19企业经营自动化数据分析服务应用设计 21](#_Toc7587)

[4.1 题目概述 21](#_Toc12597)

[4.2 所需技术和数据 22](#_Toc16414)

[4.3 国内外有关研究现状 22](#_Toc5239)

[4.4 技术/算法分析 22](#_Toc14943)

[三、 确定的初步方案设计 25](#_Toc11578)

[1、 最终选题 25](#_Toc32419)

[2、选题原因 25](#_Toc17772)

[3、技术分析 25](#_Toc21815)

[四、 下阶段工作计划 26](#_Toc13421)

[五、 文献引用 27](#_Toc23287)

[六、 总结 28](#_Toc8789)

**目录（自拟开发类根据此目录整理内容）**

[一、 小组成员信息及分工 1](#_Toc24758)

[二、 选题分析 1](#_Toc15856)

2.1 选题背景和特点分析

2.2 可行性分析

2.3 所需技术和工具分析

[三、 确定的初步方案设计 4](#_Toc3248)

[3.1 产品的基本功能描述 4](#_Toc30873)

3.2 初步需求分析

[3.2 项目的基本原型示例 7](#_Toc25916)

[四、 下阶段工作计划（具体开发过程的计划） 8](#_Toc13622)

[参考文献 9](#_Toc17178)

1. **选题分析部分**

**[1、选题一：A13歌声转换算法](#_Toc19779)**

**[1.1题目概述](#_Toc21860)**

**整体背景：**随着深度学习的发展, 近些年来音频算法在智能医疗、语音识别、语音合 成、声源定位等领域的应用非常火热,其效能和速度也不断得到精进。从过去的

云端服务,逐步脱离并发展到PC端边缘运算,到这几年再往移动端运算发展。尽管硬件的运算能力越来越强大, 但是模型轻量化仍是算法设计所追求的核心之一。唯有如此，AI算法才可以随时随地调用，发挥它们强大的作用。AIGC席卷下，语音合成，语音转换一直是音频技术的关键核心技术。语音合成的关键是学会目标人的音色，并迁移到源目标上。随着元宇宙的爆火，虚拟人逐渐走到了荧幕前，语音合成也从音色转换升级至目标人的音色克隆。即给定足够时长的目标人语音，即可通过文本的方式生成用户需要的音频片段。然而仅仅让虚拟人说话是不够的，歌声是一个很好表现自己的方式，歌声转唱就是典型的应用案例。随着“AI孙燕姿”的成功出圈，让用户可以链接粉丝和偶像产生互动感，具有很强娱乐性，除此之外，歌声转唱可以为每个人提供个性化的音乐体验，用户通过录制简单的一段说话音频，就可以定制属于自己独特音色的歌手，让不擅长唱歌的用户完成歌曲的演唱。这就是:Singing Voice Conversion。歌声转换（Singing voice conversion）旨在保证歌唱内容的同时，将音色从source speaker转换到target speaker，这将为音乐行业带来革命性的变化。歌声转唱具有很强的娱乐性，可以使目标歌手无需重新录制歌曲，不仅能够节省大量的时间和金钱，还能为艺术创作开辟新的艺术道路，同时可以让每一个用户都能实现自己独特的音色唱歌，商业前景和市场巨大。

**问题描述：**本题着力于解决歌唱声音转换问题。歌唱声音转换可以为每一个人搭建歌 唱数据集，可以学习目标人音色，使用音频相关深度算法和模型框架，完成歌曲中的音色替换。本项目要求能够设计一个在不使用GPU的环境下运行, 在精细度、速度和模型大小上取得平衡的最佳化的算法模型。此外，本项目要求实施者自行以收集开源数据或自主建立数据的方式收集数据、建立音频多人声分离算法模型。并完成模型训练、优化、工程化（或C++皆可）等工作，最终产出一个可执行程序，针对5个目标音色进行歌唱声音转换。

**[1.2所需技术和数据](#_Toc31301)**

**所需技术：**数据收集，特征提取，建立转换模型，训练和优化，音频转换。

**所需数据：**需要收集大量的说话音频和对应的歌唱音频作为训练数据。这些数据可以来自于专业歌手的录音或者公开的歌唱数据库。

**[1.3 国内外有关研究现状](#_Toc12869)**

国外方面，早在2006年，美国麻省理工学院就推出了一个基于隐马尔可夫模型（Hidden Markov Model, HMM）的歌声转换系统。之后，又有一些其他的研究者提出了基于神经网络的方法，如使用DNN来进行高质量的歌声合成。2018年，谷歌发布了WaveNet语音合成系统，该系统使用了一种基于深度神经网络的自回归模型，可以生成高质量的音频。

国内方面，2019年中国科学院计算技术研究所也推出了一款歌声转换软件，名为“星球唱将”，可以将用户的说话转换为歌声，并且支持用户自定义歌曲的编排和调整。此外，还有一些其他的国内研究团队也在进行歌声转换技术的研究，如使用深度神经网络和声码器来实现高效的歌声转换。

**[1.4 技术/算法分析](#_Toc15121)**

1. **数据收集：**首先需要收集大量的说话音频和对应的歌唱音频作为训练数据。这些数据可以来自于专业歌手的录音或者公开的歌唱数据库。
2. **特征提取：**对于每个音频样本，需要从中提取出关键的声音特征，如音高、共振峰频率、声音强度等。这些特征可以通过数字信号处理技术和声音分析算法来提取。
3. **建立模型：**使用机器学习算法，如深度神经网络（Deep Neural Networks, DNN）或循环神经网络（Recurrent Neural Networks, RNN），建立一个转换模型。该模型将输入的说话音频映射到输出的歌唱音频，同时保留原始音频的个性特征。
4. **训练和优化：**使用收集到的数据集来训练模型，并通过优化算法不断调整模型参数，以提高模型的准确性和效果。常见的的优化算法包括梯度下降法和反向传播算法。这个过程可能需要大量的计算资源和时间。
5. **音频转换：**当模型训练完成后，用户可以通过录制自己的说话音频，并使用训练好的模型将其转换为歌声。这一转换过程可以在计算机上进行，需要运行训练好的模型并对音频进行处理。

**[1.5 可行性分析与技术难点](#_Toc24633)**

**可行性分析：**

1. 数据可行性：歌声转换需要大量的音频数据作为训练集，这些数据可以来自于专业歌手的录音或者公开的歌唱数据库。这些数据在现代数字化环境下易于获取和处理。

2. 技术可行性：目前已经涌现出了许多基于深度学习的语音合成技术，歌声转换作为其中的一种，也取得了很好的效果。

3. 应用可行性：随着科技的进步，人工智能技术正在越来越广泛地应用于各个领域。从音乐产业的角度来看，歌声转换有着很大的应用前景，可以广泛应用于音乐制作、娱乐、教育等领域。

**技术难点：**

1. 数据集的质量和数量：歌声转换需要大量的训练数据，并且这些数据必须是高质量的。由于人声音色的差异性很大，所以需要收集尽可能多样化的数据，以提高模型的泛化性能和鲁棒性。

2. 特征提取算法的准确性：歌声转换需要提取出合适的音频特征作为模型的输入，这些特征必须准确地表达出音频信号中的信息。由于人声信号复杂多变，所以需要研究和改进不同的特征提取算法，以提高模型的精度和效果。

3. 模型的训练和优化：歌声转换模型需要进行大量的训练和优化，以达到较高的准确性和泛化性能。这涉及到大量的计算资源和时间，需要建立高效的训练方法和算法。

4. 音频转换的质量评估：歌声转换的质量评估是一个复杂的问题，需要针对不同的应用场景和用户需求进行评估。同时，还需要开发合适的评估指标和方法，以保证转换质量的客观性和可靠性。