**武汉大学计算机学院**

**本科生课程设计报告**

**Lemon5智能语音系统的总体设计与实现**

专 业 名 称：计算机科学与技术（弘毅班）

课 程 名 称：大型应用软件设计

团 队 名 称：Software Now

指 导 教 师：王健 副教授

团 队 成 员 一：陈胤良（2020300004031）

团 队 成 员 二：胡瑞（2020300004044）

团 队 成 员 三：王乔林（2020301052060）

团 队 成 员 四：邹奕杨（2020300004080）

团 队 成 员 五：周恋程（2020302121324）

团 队 成 员 六：谭垒鑫（2019300003096）

二○二三年十一月

**郑 重 声 明**

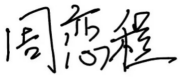
本团队呈交的设计报告，是在指导老师的指导下，独立进行实验工作所取得的成果，所有数据、图片资料真实可靠。尽我所知，除文中已经注明引用的内容外，本设计报告不包含他人享有著作权的内容。对本设计报告做出贡献的其他个人和集体，均已在文中以明确的方式标明。本设计报告的知识产权归属于培养单位。











团队成员签名： 日期：2023.11.15

摘 要

大型应用软件设计课程实验的目标是结合身边的实际需求，自选题目开发Web应用系统。

实验设计主要遵循软件工程的基本要求，采用敏捷、迭代式开发。

实验内容主要包括：需求分析、架构设计，详细设计，迭代开发、测试等。

实验结论为小组最终完成了Lemon5智能语音系统的编写，实现了一款轻量级集成性多媒体信息处理软件，降低了用户使用多媒体信息处理软件的操作成本。

**关键词：**敏捷开发；Web应用；软件工程

**目 录**

**1 实验目的和意义**

1.1 实验目的 ………………………………………………………………………1

1.2 实验意义………………………………………………1

(各章的名称黑体4号，其余宋体小4)

**……**

**……**

**……**

**2**□**实验设计**

2.1□概述……………………………………………………………………………… 35

2.2□实验原理………………………………………………………………………… 37

2.3□实验方案………………………………………………………………………… 39

**……**

**……**

**……**

**结论** …………………………………………………………………………………… 57

**参考文献 ………………………………………………………………………………** 59

**附录 ……………**………………………………………………………………………… 72

(结论、参考文献、致谢及附录黑体4号)

**1 项目选题**

**1.1** **选题背景**

现如今音乐制作、电影制作、广播电视等领域对音频处理的需求日渐强烈，个人用户对音频文件的处理和编辑也有一定的需求，故团队计划开发轻量级多媒体音频处理软件 Lemon5，降低用户使用多媒体信息处理软件的操作成本。

需要注意的是，目前市场上已经存在许多音频处理软件，包括商业软件和开源软件，例如 Adobe Audition、Ableton Live、Audacity、Reaper 等。这些软件具有强大的功能和广泛的用户群体，对于新进入市场的轻量级音频处理软件而

言，竞争压力较大。但随着数字化技术的不断发展和普及，音频处理软件的市场需求将会继续增长，对于我们来说既是机遇也是对我们能力的考验。

我们认为，未来的音频处理软件将会更多与AI技术相结合，故我们计划将人工智能技术应用于音频处理中，例如应用TTS（文本转语音）模型实现特定音色的音频生成，应用大语言模型和TTS实现语音对话等等。

**1.2** **软件功能及用户澄清**

**1．2．1** **功能一：语音合成**

语音合成功能允许用户输入一串文字并选择目标说话人后合成出一段语音。 Lemon5 借助深度学习模型 VITS 实现这一功能。用户不仅能够合成出目标语音，还能通过软件提供的选项来在一定程度上控制合成语音的速度和情感。

该功能的目标用户包括视频制作者与残障人士，尤其是视障人士。在视频制作领域，文字转语音技术可以将文本信息转换为音频信息，为视频制作者提供配音支持，视频制作者可以自主选择喜欢的音色，增加视频的趣味性；该技术也可以为视障人士提供帮助，它可以将书籍、新闻、电子邮件等文本信息转换为音频信息，帮助视障人士获得更多的自主生活能力。

**1．2．2** **功能二：语音转换**

语音转换功能允许用户输入自己的语音并选择目标说话人音色后合成出一段语音。

该功能的目标用户为视频制作者，语音转化技术可以将制作者自己音色的语音信息转换为特定音色的语音信息，从而为不具备配音技能的视频制作者提供帮助。

**1．2．3** **功能三：语音对话**

语言对话功能允许用户输入自己的语音或文字，与智能助手进行对话、提问，得到智能助手生成的语音或者文字。同样，在这个功能中我们也可以选择智能助手的音色，甚至可以通过提示词预设语音助手的性格。

该功能的目标用户为希望借助智能助手获取信息或是娱乐的人群。将语音处理与大语言模型进行结合，可以让用户有自然的交互体验。语音对话既可以是娱乐性质的对话，比如利用智能助手的特定音色进行角色扮演；也可以是学习性质的对话，我们使用了智谱AI研发的千亿参数对话模型ChatGLM，通过语音进行提问和回答可以让对话者高效地获取信息。

**1.3** **方案实施**

整体项目我们拟实现为Web应用，前端使用React框架，后端使用Python Flask框架。为了更加直观地展现功能，我们选取了知名的游戏“原神”作为语音语料库，让整体功能的展示更加清晰直观生动。

在这个项目中，为实现文本转语音功能，我们调用了原神TTS模型；为了实现语音对话，我们调用了大语言模型ChatGLM进行文字对话，并将对话结果信息利用TTS模型实现文本转语音，然后展示出来。

**2 需求分析**

**2.1** **实验目的**（黑体4号）

本实验……

（宋体小4，正文行间距固定为23磅，字符间距为标准）

.......

.......

.......

**1．1．1** **目的一：检验XXX**（黑体小4号加粗）

通过DDR IPCore 对DDR 和DDR2 SDRAM进行初始化是有分别的，由于在本次项目设计过程中实际采用的是DDR SDRAM，因此本文仅仅对前者的初始化时序进行讨论。

（宋体小4号）

**3 架构设计**

（章标题段前为0.8行、段后为0.5行）

**1.1** **实验目的**（黑体4号）

本实验……

（宋体小4，正文行间距固定为23磅，字符间距为标准）

.......

.......

.......

**1．1．1** **目的一：检验XXX**（黑体小4号加粗）

通过DDR IPCore 对DDR 和DDR2 SDRAM进行初始化是有分别的，由于在本次项目设计过程中实际采用的是DDR SDRAM，因此本文仅仅对前者的初始化时序进行讨论。

（宋体小4号）

**4 详细设计**

（章标题段前为0.8行、段后为0.5行）

**1.1** **实验目的**（黑体4号）

本实验……

（宋体小4，正文行间距固定为23磅，字符间距为标准）

.......

.......

.......

**1．1．1** **目的一：检验XXX**（黑体小4号加粗）

通过DDR IPCore 对DDR 和DDR2 SDRAM进行初始化是有分别的，由于在本次项目设计过程中实际采用的是DDR SDRAM，因此本文仅仅对前者的初始化时序进行讨论。

（宋体小4号）

**5 测试**

（章标题段前为0.8行、段后为0.5行）

**1.1** **实验目的**（黑体4号）

本实验……

（宋体小4，正文行间距固定为23磅，字符间距为标准）

.......

.......

.......

**1．1．1** **目的一：检验XXX**（黑体小4号加粗）

通过DDR IPCore 对DDR 和DDR2 SDRAM进行初始化是有分别的，由于在本次项目设计过程中实际采用的是DDR SDRAM，因此本文仅仅对前者的初始化时序进行讨论。

（宋体小4号）

**6 迭代过程**

**6.1** **第一次迭代**

**6．1．1** **迭代目标**

1. 进行UI设计，实现界面；
2. 基本完成项目架构的搭建，实现前后端交互，完善API调用以及数据的传递转换；
3. 实现登录功能；
4. 实现文本转语音功能，用户可以把文本转换成目标语音，得到不同发声者音频，并可以调整音频参数。

**6．1．2** **实现过程**

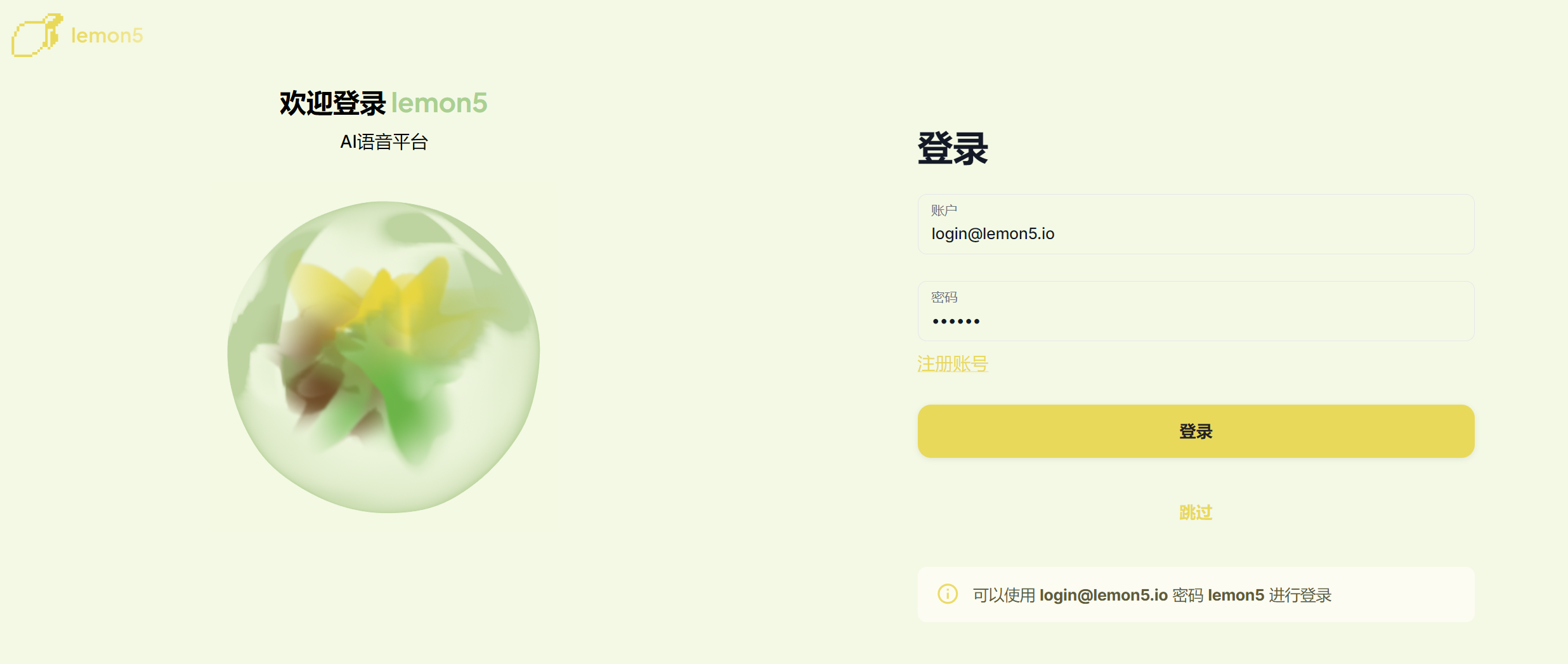
前端方面，我们设计了简洁美观的UI，并通过React框架予以实现；

后端方面，我们首先在本地部署了自己的TTS模型并进行了多轮训练，但是发现由于训练数据太少，训练结果稍微欠一点火候。于是考虑利用现成的原神TTS api，其在本地测试时具有较好效果，能够以几乎贴近原声的方式输出目标语音。

我们利用git工具完成多人协同开发，大家按照各自分工完成对应板块，再同一整合起来, 并使用sonarcube进行软件健壮性分析与错误排查。

在完成前后端交互以及api调用格式处理后，我们实现了本轮迭代全部目标。

**6．1．3** **实现结果**



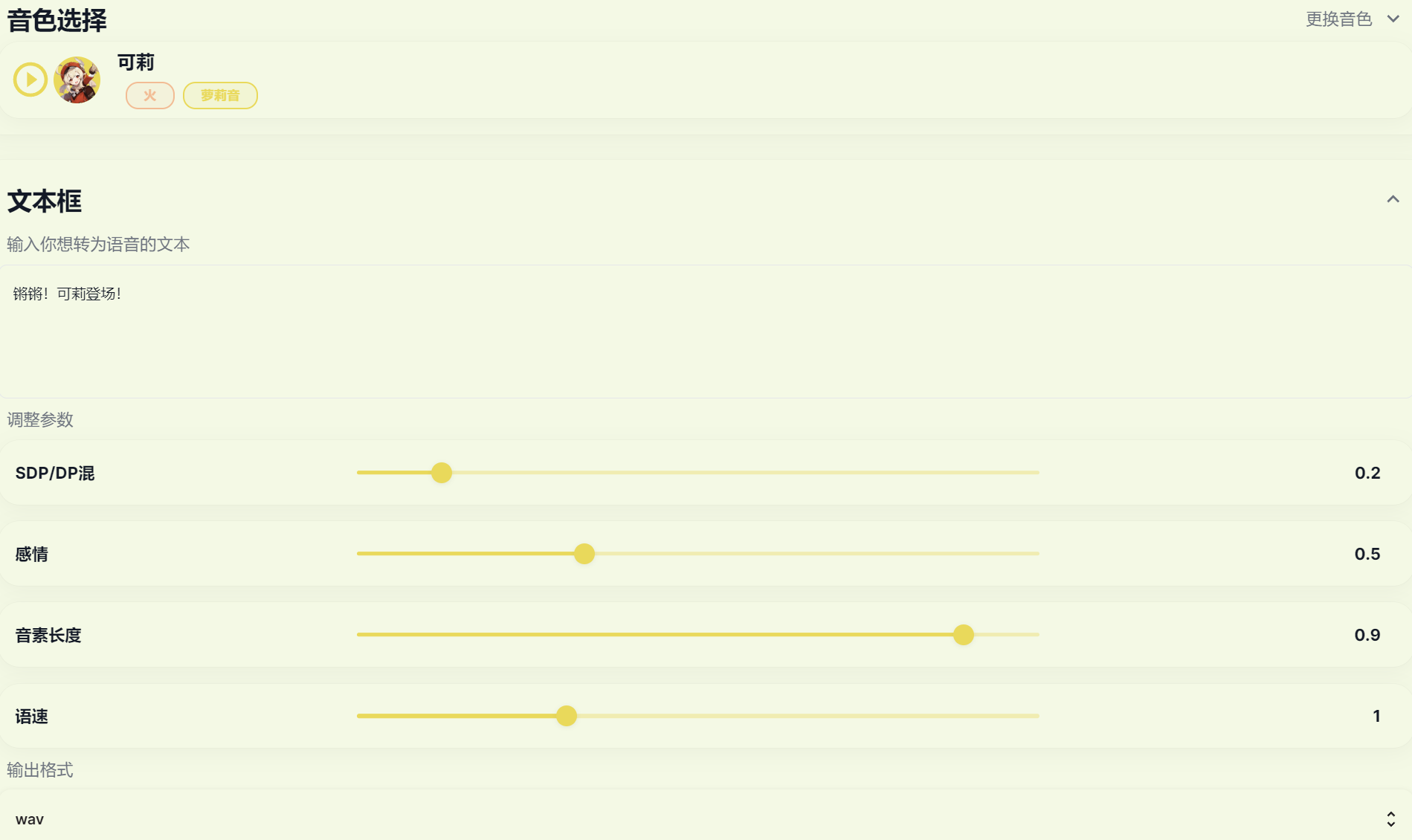
图**6.1** 登录界面



图**6.2** Lemon5主页



图**6.3** 文本转语音界面



图**6.4** 文本转语音可选项

**6.2** **第二次迭代**

**6．2．1** **迭代目标**

实现语音对话功能，用户可以通过语音与智能助手进行对话、提问，得到智能助手生成的语音或者文字。

在语音对话功能中，用户可以选择自己喜欢的音色。对于用户选择的音色，我们在后台利用一些prompt，让其具有对应人物的故事背景和性格。这样，用户就可以体验到与自己喜欢的角色沟通的乐趣。

**6．2．2** **实现过程**

在语音对话实现的过程中，我们调用国内现成大模型ChatGLM得到对话结果信息并以json格式截取，利用TTS模型实现文本转语音，并放到voicecard里面展示。

**6．2．3** **实现结果**



图**6.5** 语音对话界面

可以看出，语音对话具有不错的效果，模型能够贴切且流畅地输出目标文字语音，无障碍进行沟通。同时，我们使用的prompt效果较好，AI成功扮演了对应的角色。

**7 项目亮点**

**7.1** **亮点一：可选音色丰富**

在Lemon5中，用户可以选择目标说话人的音色来合成出一段语音，还可以通过软件提供的选项来在一定程度上控制合成语音的速度和情感，从而增加语音的表现力，语音合成功能相当灵活多变。目前我们提供了原神角色的语音和丁真的语音，日后还可以根据用户的需求提供其他音色以供选择。用户可以根据自己不同的需要，合成不同音色、情绪的语音，具有相当的实用性和趣味性。

例如，用户可以输入自己的语音，说“你好，我的动物朋友雪豹”，然后选择丁真的音色，Lemon5 就会合成出一段丁真的语音，让用户感受到丁真的亲切和幽默。

**7.2** **亮点二：娱乐性**

Lemon5具有智能对话的功能，用户可以用自己的语音或文字与 AI 聊天。在这个功能中，用户可以与流行的游戏作品《原神》中的人物对话，或是与网红丁真对话。这样，用户就可以体验到与自己喜欢的角色或者明星沟通的乐趣。

例如，用户可以输入自己的语音或文字，说“你好派蒙，你是应急食品吗”，然后选择原神中的派蒙，Lemon5 就会用 AI 聊天的方式回复用户，说“是的，我是应急食品”，让用户感受到与原神中人物对话的趣味。

**7.3** **亮点三：模型性能优秀**

Lemon5的亮点之一是使用了性能优秀的VITS模型。VITS是一个用于端到端文本转语音的模型，它可以同时学习语音的内容、风格和声码器，从而实现更自然和更灵活的语音合成效果。VITS还结合了变分推理、标准化流和对抗训练，这些技术可以提高语音的质量和多样性，也可以减少对训练数据的需求，这意味着我们可以通过更少的训练样本达到更好地语音效果，有利于未来我们训练更多音色以供选择。

**8 团队成员贡献**

（章标题段前为0.8行、段后为0.5行）

**1.1** **实验目的**（黑体4号）

本实验……

（宋体小4，正文行间距固定为23磅，字符间距为标准）

.......

.......

.......

**1．1．1** **目的一：检验XXX**（黑体小4号加粗）

通过DDR IPCore 对DDR 和DDR2 SDRAM进行初始化是有分别的，由于在本次项目设计过程中实际采用的是DDR SDRAM，因此本文仅仅对前者的初始化时序进行讨论。

（宋体小4号）

**9 成员个人总结**

（章标题段前为0.8行、段后为0.5行）

**1.1** **实验目的**（黑体4号）

本实验……

（宋体小4，正文行间距固定为23磅，字符间距为标准）

.......

.......

.......

**1．1．1** **目的一：检验XXX**（黑体小4号加粗）

通过DDR IPCore 对DDR 和DDR2 SDRAM进行初始化是有分别的，由于在本次项目设计过程中实际采用的是DDR SDRAM，因此本文仅仅对前者的初始化时序进行讨论。

（宋体小4号）

**参考文献**

[1] 戴军，袁惠新.膜技术在含油废水处理中的应用[J].膜科学与技术，2002，22（2）：59-64.

[2] 毛侠，孙云.和谐图案的自动生成研究[A].第一届中国情感计算及智能交互学术会议论文集[C].北京：中国科学院自动化研究所，2003：277-279.

[3] 王湛.膜分离技术基础[M].北京:化学工业出版社，2000：14-21，30.

[4] 张志祥. 间断动力系统的随机扰动及其在守恒律方程中的应用[D].北京:北京大学数学学院,1998.

[5] World Health Organization. Factors regulating the immune response: report of WHO Scientific Group[R]. Geneva: WHO, 1970.

[6] 河北绿洲生态环境科技有限公司.一种荒漠化地区生态植被综合培育种植方法:中国,01129210.5[P].2001-10-24.

[7] GB/T16159-1996,汉语拼音证词法基本规则[S].北京：中国标准出版社，1996.

[8] 毛侠.情感工学破解“舒服之谜”[N].光明日报，2004-04-17（B1）.

[9] 陈剑.上博简《民之父母》“而得既塞於四海矣”句解释[EB/OL］.简帛研究网站，http://www.bamboosilk.org/Wssf/2003/chenjian03.htm．2003-01-18.