

项目编号： 2024PY—156

2024 年度本科生创新创业训练培育项目
中期汇报表

项目名称		国际中文朗读语料评分模型 建构与优化实践
项目 负责人	姓名	曹雅珈
	院系	国际汉语文化学院
	专业	汉语国际教育
	手机	13910389830
	邮箱	10222740425@stu. ecnu. edu. cn
指导教师姓名、职称		凌锋副教授
项目立项时间		2024

填表日期：2024 年 11 月 9 日

项目 基本情 况	项目选题来源 (自立项目或教师科研课题的子项目)	教师科研课题的子项目
	依托单位或合作单位 (实验室、研究中心、中小学机构等)	无

一、项目计划达到的目标和内容

该项目计划的核心目标是通过改进评分标准、接入 AI 大模型和功能拓展来探索更加科学的中文朗读评分标准，并初步打造一个兼具科学性、易用性和高效性的中文朗读评分及智能学习平台，帮助学生提升语言水平并为教师提供更优质的教学工具。

具体来说，在完善评分标准方面，我们将注重简化和优化评分方法，使其在模型训练与实际应用中更加平衡。现有的评分标准由于过于细致，在实际操作中并不易用。为了让评分更贴近人类评估习惯，我们计划调整为以音节为单位进行综合评分，减少声母、韵母和声调的细分，以更好地反映评分者的整体印象。与此同时，我们将改进模型对评分宽严度的把控，确保评分体系既能满足模型训练需求，又便于实际应用操作。此外，为应对学生在集体学习中的相对评价需求，我们将基于合理的分布调整分数计算方式，使评分体系更能反映学生的真实水平并适应不同学习背景。

在模型训练方面，我们计划构建一个切分准确、评分合理、科学量化的大规模语料数据库。所有语音都是根据国际中文学习者在学习汉语的过程中朗读的真实语料制作而成。计划构建的数据库中包含不少于 5000 条可供模型学习的有效音频，总时长超过 3000 分钟，每条数据由熟练掌握语音学知识、经过系统性培训的专业人员处理评分，可以很好地应用于 AI 大模型的训练。

在平台功能拓展方面，我们重点聚焦学生端和教师端的完善与优化，计划应用智谱 AI 平台的 GLM-4-9B 模型，并采用 LoRA (Low-Rank Adaptation) 微调技术，对多项模型训练参数进行调整和优化，辅助以 Java 编写后端 API 和讯飞语音评测平台功能，初步建立一个多功能、智能化的在线学习平台，旨在为有需求的中文学习者提供高效、便捷的口语学习和朗读练习网页。预计建立的评分模型

准确率不低于 85%，F1 分数不低于 80%。

在学生端方面，我们将分阶段实现平台功能优化。第一阶段将重点实现以下功能：引入《发展汉语》等权威教材的多样化朗读材料，确保学习内容在难度、主题和趣味性上丰富多样，帮助学生根据个人水平进行有针对性的练习。此外，我们预计开发“进步曲线图”等功能，使学生能够直观地查看学习成果，并通过每日签到和积分奖励系统保持学习动力。我们还将在这一阶段提供标准音频和舌位动画，帮助学生模仿和纠正发音，提供评分之外的额外指导性意见。第二阶段将引入更多互动和良性竞争性功能，例如“每日阅读比拼”模块。学生可以与其他学习者进行朗读成绩竞争，实时动态榜单将激励他们不断练习，以确认自己的学习水平并保持积极的学习氛围。我们将持续收集和分析用户数据以调试系统，做到模型优化贯穿两大阶段，提升评分的准确性和自动化分析能力，从而实现更合理的评分。

在教师端的建设上，我们将分阶段推进平台功能开发。初期将实现朗读作业的布置与检查功能，教师可以设定朗读任务并批量查看学生提交的录音。自动评分功能将作为模块上线，帮助教师根据系统初步评分进行调整，减轻其评分负担。第一阶段结束后，我们会收集使用反馈并进行功能优化。录音文件管理功能将在第二阶段上线，教师可以批量管理和分析学生录音，并通过分类、标签管理和搜索筛选功能快速定位重点内容。这些功能将在收集反馈后进行进一步完善。未来，我们计划通过数据积累和分析，逐步实现大模型支持，开发自动化教学建议功能，帮助教师识别和纠正学生的发音问题。初步的个性化建议功能将在大模型上线后进行测试和优化。

我们计划在该平台上线后将开发出的大模型网页提供给学院内国际中文师生先行开展产品内测，进入我院语言课堂（语言班 1-4D），在不干涉教师正常教学进度的情况下推广平台，丰富数据并结合内测反馈及时改进和完善现有模型；在初步完善的基础上，继续扩大测试规模，预计平台注册量不少于 500 人次，并对项目进行多平台推广，邀请多学段、多层次的汉语学习者进行产品体验，有效推进产品普适性的提升。

二、研究进展和当前成果

经过前期努力，本项目目前已取得系列阶段性成果，以下将从数据处理、评分制定和网页呈现等方面分别进行介绍。

1、文本与音频数据收集

我们已处理的录音文件数据来源于学院自建的韩国留学生语料库，现有 5000 余条可供模型学习的有效音频，总时将近 3000 分钟，其相应文本出自于北大版《高级汉语听说教程》以及《中级汉语听说教程》上册教材。朗读文本以对话或独白为主，辅以一些生词的朗读内容。

经多次文本分段、数据调试与云端模型运行，我们最终确定了单个音频上传时长，在克服讯飞模型的限制的同时最大化评分效率。我们将课文按照话轮或内在逻辑关系进行人工分段，并将音频文件归一化处理为 16000hz 文件，方便读取使用；根据分段结果自动剪切已有音频文件并重新编号；人工复检分段是否有误。

2、讯飞音频分析

在本阶段研究中，我们主要采用开源的讯飞评分模型进行音频分析，并通过 Python 实现模型接入，同时利用 Praat 进行更细致的语音分析。该过程利用系列程序与 AI 大模型，实现了语音文本数据格式处理和识别转换。我们编写完成了 audio.py 脚本，使预处理后的音频文件能够自动地置入指定文件夹并生成 pcm 文件，实现了讯飞模型的音频识别功能。随后，我们编写了 ise_ws_python3_demo.py 脚本，实现了 xml 数据文件的生成。在此基础上，我们通过编写 transfer_score.py 代码实现了 xlex 和 txt 文件的生成，进而完成了讯飞评测的接入过程。我们还编写了 Praat 调用脚本，实现了 textgrid 文件的生成功能，以便进行语音标注和分析。（见图 1、图 2、图 3）。

```

9 # 检查子目录是否存在, 如果不存在则创建
10 if not os.path.exists(subdirectory):
11     os.makedirs(subdirectory)
12
13 # 遍历子目录中的所有音频文件 (.wav 和 .mp3)
14 for filename in os.listdir(subdirectory):
15     if filename.endswith((''.wav', '.mp3')): # 检查文件扩展名
16         # 构建完整的文件路径
17         file_path = os.path.join(subdirectory, filename)
18         print(f"Processing {file_path}...") # 打印当前正在处理的文件名
19
20         # 读取音频文件
21         y, sr = librosa.load(file_path, sr=None, mono=True) # 自动获取
22         print(f"Original sample rate: {sr}") # 打印原始采样率
23
24         # 重新采样到16000 Hz
25         y_16 = librosa.resample(y, orig_sr=sr, target_sr=16000)
26         print(f"Resampled to 16000 Hz")
27
28         # 保存处理后的文件为16000Hz的WAV文件
29         new_wav_file_path = os.path.splitext(file_path)[0] + '_16000.wav'
30         sf.write(new_wav_file_path, y_16, 16000, format='WAV')
31         print(f"Written WAV file at 16000 Hz: {new_wav_file_path}")
32
33         # 如果是MP3文件, 处理后删除原始文件
34         if filename.endswith('.mp3'):
35             os.remove(file_path)
36             print(f"Deleted original MP3 file: {file_path}")
37
38         # 转换为PCM格式, 裁剪到[-1.0, 1.0]范围内
39         y_pcm = array.array('h', (int(max(min(sample, 1.0), -1.0) * 32768) for sample in y_16))
40         output_file = os.path.splitext(new_wav_file_path)[0] + '.pcm'
41         with open(output_file, 'wb') as pcm_file:
42             y_pcm.tofile(pcm_file)
43         print(f"Written PCM file: {output_file}")
44
45 def read_audio_text_mapping():
46     with open('cn/音频-文本对照表.csv', mode='r', encoding='utf-8') as csv_file:
47         reader = csv.reader(csv_file)
48         next(reader) # Skip header
49         for row in reader:
50             audio_text_mapping[row[0]] = row[1]
51     return audio_text_mapping
52
53 class Ws_Param(object):
54     # 初始化
55     def __init__(self, APPID, APIKey, APISecret, AudioFile, Text):
56         self.APPID = APPID
57         self.APIKey = APIKey
58         self.APISecret = APISecret
59         self.AudioFile = AudioFile
60         self.Text = Text
61
62     # 公共参数(common)
63     self.CommonArgs = {"app_id": self.APPID}
64     # 业务参数(business), 更多个性化参数可在官网查看
65     self.BusinessArgs = {"category": CATEGORY, "sub": SUB, "ent": ENT, "cmd": "ssb", "auf": "aue", "raw": "raw", "text": self.Text, "http_skip": True, "aus": 1}
66
67     # 生成url
68     def create_url(self):
69         url = 'ws://ise-api.xfyun.cn/v2/open-ise'
70         now = datetime.now()
71         date = format_date_time(mktime(now.timetuple()))
72
73         signature_origin = "host: " + "ise-api.xfyun.cn" + "\n"
74         signature_origin += "date: " + date + "\n"
75         signature_origin += "GET " + "/v2/open-ise" + " HTTP/1.1"
76         signature_sha = hmac.new(self.APISecret.encode('utf-8'), signature_origin.encode('utf-8'), digestmod=hashlib.sha256).digest()
77         signature_sha = base64.b64encode(signature_sha).decode(encoding='utf-8')
78
79         authorization_origin = "api_key=\"%s\", algorithm=\"%s\", headers=\"%s\", signature=\"%s\"" % (self.APIKey, "SHA256", "host, date, GET", signature_sha)

```

图 1 讯飞识别音频代码展示（部分）

```

def on_open(ws):
    def run(*args):
        with open(wsParam.AudioFile, "rb") as fp:
            while True:
                # 发送最后一帧
                d = {"business": {"cmd": "auw", "aus": 4, "aue": "raw"},
                    "data": {"status": 2, "data": str(base64.b64encode(segment[st
                    time.sleep(1)

                # 更新已匹配的字索引
                current_char_index += len(segment_text)

            ws.close()

        thread.start_new_thread(run, ())

if __name__ == "__main__":
    audio_text_mapping = read_audio_text_mapping()

    for audio_file, text in audio_text_mapping.items():
        # 尝试读取文件
        audio_file_open = audio_file + '_16000' # 添加后缀
        audio_file_path = os.path.join('cn', audio_file_open + '.pcm') # 构建完整
        # 处理文件的逻辑
        wsParam = Ws_Param(APPID='4884d8d1', APISecret='MG10ThkM2Q1Hjc1VzcX001IMT
        APIKey='269cba2b9e41326b0303e1a1b36b208b',
        AudioFile=os.path.join('cn', audio_file_open + '.pcm')

        if os.path.exists(audio_file_path):
            websocket.enableTrace(False)
            wsUrl = wsParam.create_url()
            ws = websocket.WebSocketApp(wsUrl, on_message=on_message, on_error=on_
            ws.on_open = on_open
            ws.run_forever(sslopt={"cert_reqs": ssl.CERT_NONE})

    if filename.endswith(".xml"):
        # 创建空的DataFrame来存储当前文件的数据
        all_data = []

        # 解析 XML 文件
        tree = ET.parse(os.path.join("cn", filename))
        root = tree.getroot()

        # 遍历 XML 树
        for read_sentence in root.findall('.//read_sentence'):
            for word in read_sentence.findall('.//word'):
                sylls = word.findall('.//syll')
                for syll in sylls:
                    syll_data = {
                        'filename': filename,
                        'content': syll.get('content'),
                        'beg_pos': syll.get('beg_pos') if syll.get('beg_pos') else "-1",
                        'end_pos': syll.get('end_pos') if syll.get('end_pos') else "-1",
                        'symbol': syll.get('symbol'),
                        'time_len': syll.get('time_len'),
                        'rec_node_type': syll.get('rec_node_type')
                    }
                    all_data.append(syll_data)
                phones = syll.findall('.//phone')
                for phone in phones:
                    phone_data = {
                        'filename': filename,
                        'content': phone.get('content'),
                        'beg_pos': phone.get('beg_pos') if phone.get('beg_pos') else "-1",
                        'end_pos': phone.get('end_pos') if phone.get('end_pos') else "-1",
                        'symbol': phone.get('symbol'),
                        'time_len': phone.get('time_len'),
                        'rec_node_type': phone.get('rec_node_type'),
                        'dp_message': phone.get('dp_message'),

```

图 2 讯飞测评接入代码展示（部分）

```

24 ;读取表格中的数据
25 for iFile to fileTotal
26   selectObject: "Strings Score"
27   seq$ = Get string: iFile
28   seq$ = seq$ - ".txt"
29   seq$ = replace$(seq$, "(", " ", 0)
30   seq$ = replace$(seq$, ")", " ", 0)
31   selectObject: "Table 'seq$'"
32   rowTotal = Get number of rows
33   beg_pos# = zero#(rowTotal)
34   end_pos# = zero#(rowTotal)
35   time_len# = zero#(rowTotal)
36   symbol# = zero#(rowTotal)
37   is_yun# = zero#(rowTotal)
38   rec_node_type# = zero#(rowTotal)
39   dp_message# = zero#(rowTotal)
40   perr_msg# = zero#(rowTotal)
41   for i to rowTotal
42     content$ = Get value: i, "content"
43     beg_pos = Get value: i, "beg_pos"
44     end_pos = Get value: i, "end_pos"
45     symbol$ = Get value: i, "symbol"
46     time_len = Get value: i, "time_len"
47     rec_node_type# = Get value: i, "rec"
48     is_yun = Get value: i, "is_yun"
49     dp_message = Get value: i, "dp_mess"
50     perr_msg = Get value: i, "perr_msg"
51     beg_pos[i] = beg_pos * 0.01
52     end_pos[i] = end_pos * 0.01
53     time_len[i] = time_len * 0.01
54     recLen = length(rec_node_type$)
55     rec_node_type[i] = recLen
56     is_yun[i] = is_yun
57     dp_message[i] = dp_message
58     perr_msg[i] = perr_msg
59     symLen = length(symbol$)
60     if symLen > 1
67 ;获得有效数据的开始阶段,这是为了排除最开始
68 break = 0
69 for m to rowTotal
70   if symbol#[m] = 1
71     break = break + 1
72     if break = 1
73       first = m
74     endif
75   endif
76 endfor
77 ;插入边界
78 selectObject: "Sound 'seq$'"
79 To TextGrid: "syl pinyin shengyun s"
80 for j from first to rowTotal
81   count = 0
82   selectObject: "TextGrid 'seq$'"
83   beg_exi = Get interval boundary
84   end_exi = Get interval boundary
85   if rec_node_type#[j] = 5
86     if symbol#[j] = 1
87       count = 1
88     endif
89   elseif rec_node_type#[j] = 3
90     if dp_message#[j] >= 0
91       count = 1
92     endif
93   endif
94   if count = 1
95     if beg_exi = 0
96       if beg_pos#[j] > 0
97         Insert boundary: 1, 200
98         Insert boundary: 2, 201
99         Insert boundary: 3, 202
100       . . . . .
128 ;填入信息
129 intSV = 2
130 for q from first to rowTotal
131   count = 0
132   selectObject: "Table 'seq$'"
133   content$ = Get value: q, "content"
134   if is_yun#[q] < 2
135     if end_pos#[q] > 0
136       if beg_pos#[q] > 0
137         count = 1
138       endif
139     endif
140   endif
141   if rec_node_type#[q] = 3
142     if dp_message#[q] >= 0
143       count = 1
144     endif
145   endif
146   if count = 1
147     selectObject: "TextGrid 'seq$'"
148     Set interval text: 3, intSV, content$
149     if perr_msg#[q] = 1
150       Set interval text: 5, intSV, "声韵错"
151     elseif perr_msg#[q] = 2
152       Set interval text: 5, intSV, "调型错"
153     elseif perr_msg#[q] = 3
154       Set interval text: 5, intSV, "声韵调型错"
155     endif
156     intSV = intSV + 1
157   endif
158   selectObject: "TextGrid 'seq$'"
159   Save as text file: defaultDirectory$ + "\textgrid\"s
160   selectObject: "Sound 'seq$'"
161   Save as WAV file: defaultDirectory$ + "\textgrid\"se
162 endfor
163 ;select all

```

图 3 praat 脚本调用代码展示（部分）

讯飞分析结果可提供音韵错、声调错、漏读、误读、增读、回读等错误类型，但在声韵错误方面未明确指出具体是声母还是韵母问题，也未提供详细的指正。因此，基于错误分析理论（Error Analysis Theory）和跨语言研究和语言迁移理论，我们在讯飞分析基础上进行了人工标注，详细记录声母和韵母的具体偏误，更精确地定位错误的来源，以揭示学生在中文学习中普遍存在的“共性”错误与“个性”错误。标注格式统一为如“/p/->/b/”表示送气双唇塞音误读为不送气双唇塞音（语内偏误）；“[ɕ]->[ʃ]”用于标识语际偏误。韵母标注方式类似。标注完成后，我们导出 excel 表格进行错误类型分析，验证了该标注方法的清晰性与合理性，清晰地呈现学生常见的错误类型。

在声调分析方面，讯飞未能识别调型问题。因此，我们在其评注基础上补充标注了声调错误类型，主要分为声调混读和怪调。声调混读如“一声->二声”，表明学生将一声误读为二声，反映出其对声调的认知或记忆偏差；怪调则指学生使用了四声之外的声调，如“一声->低升”表示一声误读为低升。textgrid 文件的标注包括：第一行为原字，第二行为正确音节，第三行为正确声韵，第四行为回读与漏读，第五行为错误声韵类型，第六行为错误声调类型，最后一行为其他标注，如音长问题和儿化音错误（见图 4）。

此外，由于讯飞模型在评分宽严度上存在局限性，如部分音节评分过严或过松，特别是在声调评分方面，我们通过人工标注进一步优化了评分结果和评分依据。

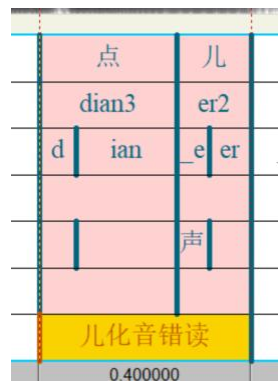


图 4 儿化音错误 textgrid 文件标注示例图

3、评分标准制定与人工评分

在本阶段，我们已初步完成了评分体系的构建，确保其在模型训练中的应用具有充分的差异性和多维度覆盖。具体而言，我们已结合机器评分与人工评分建立了综合评价框架。机器评分部分由讯飞提供的评分结果和基于其反馈信息计算的分数组成，涵盖多重维度，相互补充。人工评分方面，我们已落实了对机器评分细节的补充，包括声母、韵母和声调的逐项评估，同时对情感维度的评分方法进行了初步尝试，但尚未完全落地。

在数据处理方面，我们已经应用了 Z-Score 和 T 值法对数据进行了量化和归一化处理，生成了百分制的评分结果。此评分目前用于对模型评分标准的调整和优化，以确定更为合理的权重和评分体系，并为后续反馈机制提供依据。

截至目前，我们已人工评分语料 2000 余条，涵盖准确度、流利度、完整度等 9 大特征维度。准确度方面，涵盖了声母、韵母和声调的细化评分；流利度方面，包含了语速和停顿的评估。在评分完成后，我们会提供人工评语，并结合机器评分结果，形成更为合理的评分逻辑。通过模型反馈，我们进一步调整了数据权重和评分标准，逐步完善整个评分体系。

4、大模型训练

(1) 模型选择

考虑到数据量较少、并发需求低、计算资源和资金有限，我们选择了智谱 AI 平台的 GLM-4-9B 模型（图 5）以简化部署和维护，并采用 LoRA 微调技术，以应对中小数据集上的中文口语评分任务，强调泛化和鲁棒性，模型参数配置根据验证数据和损失函数表现进行优化。

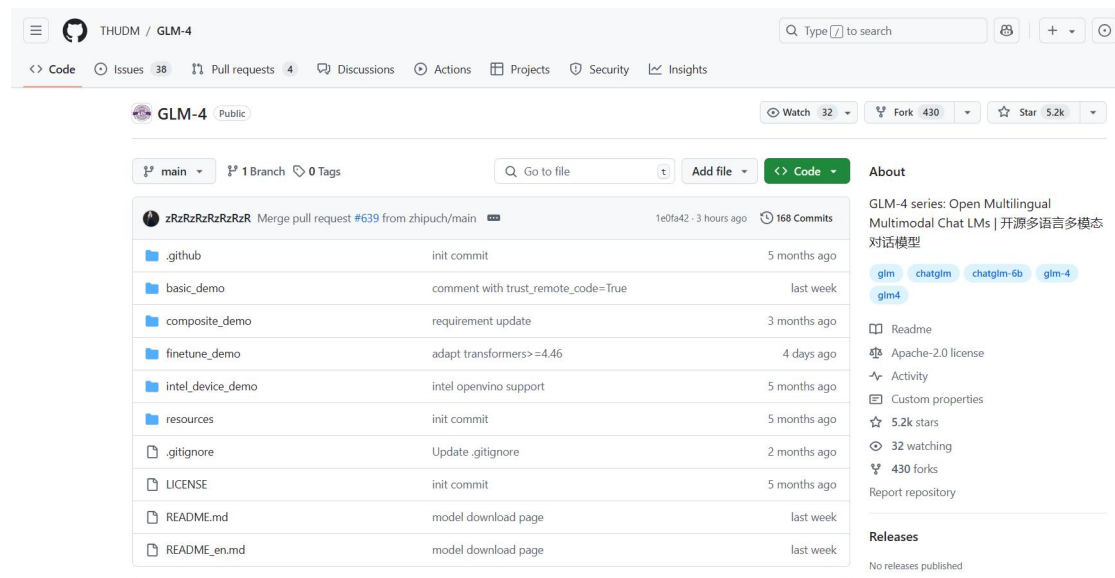


图 5 智谱 AI GLM-4 模型

(2) 模型训练方式

为了使模型更好地理解任务，我们精心设计了训练数据结构。训练数据（图 6、图 7）中包含了学生的朗读文本、音频量化数据以及评分标准，确保模型能够准确捕捉到评分的关键信息。LoRA 微调技术可以在模型权重矩阵中引入低秩适应矩阵，可以在不影响模型性能的前提下，大幅度减少训练参数的数量。

A	B	C	D	E	F	G	H	I	J	K	
1	序号编号	原编号	正文内容	语音数据	语速	停顿	声谷	韵母	声母	完整度	评论
2	1-3	2-2-2	俗话说，“男大当婚，女大当嫁”，而婚姻应该建立在感情的基础之上，每个人都带自己选择恋人的标准，让我们听听他们是怎么说的吧。	此数据为留学生音频量化数据，每步数据都是口语音频中的音频信息，每步数据格式为： [content: [beg_pos, end_pos, dp_message, score, tone, is_yun, perr_arg], (数据为空格代表数据无此属性)，参数说明：['content': '音频内容', 'beg_pos': '开始边界时间', 'end_pos': '结束边界时间', 'dp_message': '0正常；16静音；32静音；64静音；128静音（当dp_message不为空时，perr_arg可超过dp_message值保持一致的情况），'score_tone': '音调', 'is_yun': '0静音；1韵母', 'perr_arg': '当is_yun=1时，perr_arg有四种状态：0静音正确；1静音错误；2音调错误；3韵母和音调均错误'（注：content为11表明是英文文本中的内容），其中给出的第一条数据是可参考的讯飞AI数据评分，量化数据如下：[1段数据分:88.141500,完整度分:94.339622,声谷分:79.245285,音调分:86.792465,总分【模型回归】:85.048126],sil:[0.62,0,...],ur:[88.98,0,T00]	2	3	3	3	3	训练不够流利，有少许声母韵母错误，但完成度较好，希望继续保持！	
3	1-4	2-2-3	（王强，男，24岁，工人。） 这几年，不断地有人给我介绍对象。因为我这人特别的讨人喜欢，就是诚实、能干，要说找对象吧，我的要求不是很高，长得好看，性格好，主要是心要好，脾气不能太火爆，比较会生活，会持家，当然啦，我也只有这一个独生子，我将来的那位妻本孝顺父母，不能惹老人生气，可是现在，就找这条件要找到合适的姑娘也不是太容易吧。	此数据为留学生音频量化数据，每步数据都是口语音频中的音频信息，每步数据格式为： [content: [beg_pos, end_pos, dp_message, score, tone, is_yun, perr_arg], (数据为空格代表数据无此属性)，参数说明：['content': '音频内容', 'beg_pos': '开始边界时间', 'end_pos': '结束边界时间', 'dp_message': '0正常；16静音；32静音；64静音；128静音（当dp_message不为空时，perr_arg可超过dp_message值保持一致的情况），'score_tone': '音调', 'is_yun': '0静音；1韵母', 'perr_arg': '当is_yun=1时，perr_arg有四种状态：0静音正确；1静音错误；2音调错误；3韵母和音调均错误'（注：content为11表明是英文文本中的内容），其中给出的第一条数据是可参考的讯飞AI数据评分，量化数据如下：[1段数据分:80.919107,完整度分:99.300997,声谷分:81.319443,音调分:83.333328,总分【模型回归】:72.493846],sil:[0.192,0,...],ur:[92.214,0,0,0],score:[214.232,0,T00],sil:[0.232,256,0,0,0],lang:[256,272,0,T00],sil:[0.411,1277,278,32,...],ur:[278,281,32,0]	4	4	4	4	4	训练较为流利，完整度高，声母韵母发音准确，声母上有些问题，优秀！	
4	1-5	2-2-4	小丽，女，20岁，重庆人 我现在还没有男朋友，为什么呢？没看上啊，不是我瞧不上他，就是性格不合，现在的小伙子，又要长相长得漂亮，还希望人家要贤惠，要会做饭，最好还有个好职业。	此数据为留学生音频量化数据，每步数据都是口语音频中的音频信息，每步数据格式为： [content: [beg_pos, end_pos, dp_message, score, tone, is_yun, perr_arg], (数据为空格代表数据无此属性)，参数说明：['content': '音频内容', 'beg_pos': '开始边界时间', 'end_pos': '结束边界时间', 'dp_message': '0正常；16静音；32静音；64静音；128静音（当dp_message不为空时，perr_arg可超过dp_message值保持一致的情况），'score_tone': '音调', 'is_yun': '0静音；1韵母', 'perr_arg': '当is_yun=1时，perr_arg有四种状态：0静音正确；1静音错误；2音调错误；3韵母和音调均错误'（注：content为11表明是英文文本中的内容），其中给出的第一条数据是可参考的讯飞AI数据评分，量化数据如下：[1段数据分:74.801172,完整度分:98.411179,声谷分:88.726026,音调分:76.003216,总分【模型回归】:74.801172]	4	3	3	4	3	主要问题集中在声母，声母韵母有少许问题，但完成度尚好，有自我纠正意识，很棒！	

图 6 用于模型训练的数据（部分）

```
# 生成JSON数据
1 个用法  RedamancyXun
def generate_json(df, row_index, scoring_criteria):
    json_data = {
        "messages": [
            {
                "role": "system",
                "content": "你是一名留学生中文口语打分的国汉老师，请对学生的口语表现进行评估。你的评分标准基于以下六个维度：语速、停顿、声母、韵母、声调和完整度，每个维度给出0-5的整数评分。",
            },
            {
                "role": "user",
                "content": "我是一位学习中文口语的留学生，我的朗读内容是：{' + str(df.iloc[row_index, 2]) + '}'。我的音频的量化数据是：{' + str(df.iloc[row_index, 3]) + '}'。",
            },
            {
                "role": "assistant",
                "content": "语速： " + str(df.iloc[row_index, 4]) + "； 停顿： " + str(df.iloc[row_index, 5]) + "； 声母： " + str(df.iloc[row_index, 6]) + "； 韵母： " + str(df.iloc[row_index, 7]) + "； 声调： " + str(df.iloc[row_index, 8]) + "； 完整度： " + str(df.iloc[row_index, 9]) + "。",
            },
        ],
    }
    return json_data
```

图 7 训练数据 json 格式生成

(3) 模型表现评估

在模型训练完成后，我们对模型在验证集上的表现进行了评估。在验证集上，模型的准确率为 85%，F1 Score 为 82%。损失函数在训练过程中逐渐收敛，表明模型能够有效地学习到评分标准和音频数据之间的关系。

(4) 模型部署

在模型部署阶段，我们通过 Java 后端 API（图 8）实现音频和文本数据的在线调用与评分，并结合讯飞中文语音识别技术（图 9），提升模型评分的准确性与稳定性，确保高质量音频数据处理。

```
// Request body
JsonObject requestBody = new JsonObject();
requestBody.addProperty( property: "model", value: "glm-4-flash");

JsonObject systemMessage = new JsonObject();
systemMessage.addProperty( property: "role", value: "system");
systemMessage.addProperty( property: "content", system);

JsonObject userMessage = new JsonObject();
userMessage.addProperty( property: "role", value: "user");
userMessage.addProperty( property: "content", prompt);

requestBody.add( property: "messages", new Gson().toJsonTree(new JsonObject[]{systemMessage, userMessage}));

// Create OkHttp client
OkHttpClient client = new OkHttpClient.Builder()
    .connectTimeout( timeout: 1000, TimeUnit.SECONDS) // 连接超时时间
    .readTimeout( timeout: 3000, TimeUnit.SECONDS) // 读取超时时间
    .writeTimeout( timeout: 1005, TimeUnit.SECONDS) // 写入超时时间
    .build();

// Create HTTP request
Request request = new Request.Builder()
    .url(url)
    .addHeader( name: "Authorization", value: "Bearer " + apiKey)
    .post(RequestBody.create(MediaType.parse( string: "application/json"), requestBody.toString()))
    .build();

// Execute request
String responseBody = null;
String responseAdvice = null;
try (Response response = client.newCall(request).execute()) {
    if (response.isSuccessful()) {
        if (response.body() != null) {
```

图 8 Java 后端 API 调用



图 9 讯飞语音测评平台

5、UI 设计与用户体验

为便利学生获取朗读评分学练资源，我们设计了一个界面简洁、易于使用的网页平台，即“即享朗读评分”。该平台旨在为用户提供中文朗读语料及配套的朗读评分工具，以支持学生的练习需求。考虑到学生在零碎时间进行口语练习的需求，我们最终选择移动端（手机）界面作为主要展示界面（图 10），实现即时录音和上传功能。此外，我们设计的吉祥物“小吉祥”会在学生上传朗读录音后为用户评分，并提供相应的评语和建议，寓教于乐，帮助学生更有针对性地提升朗读水平。

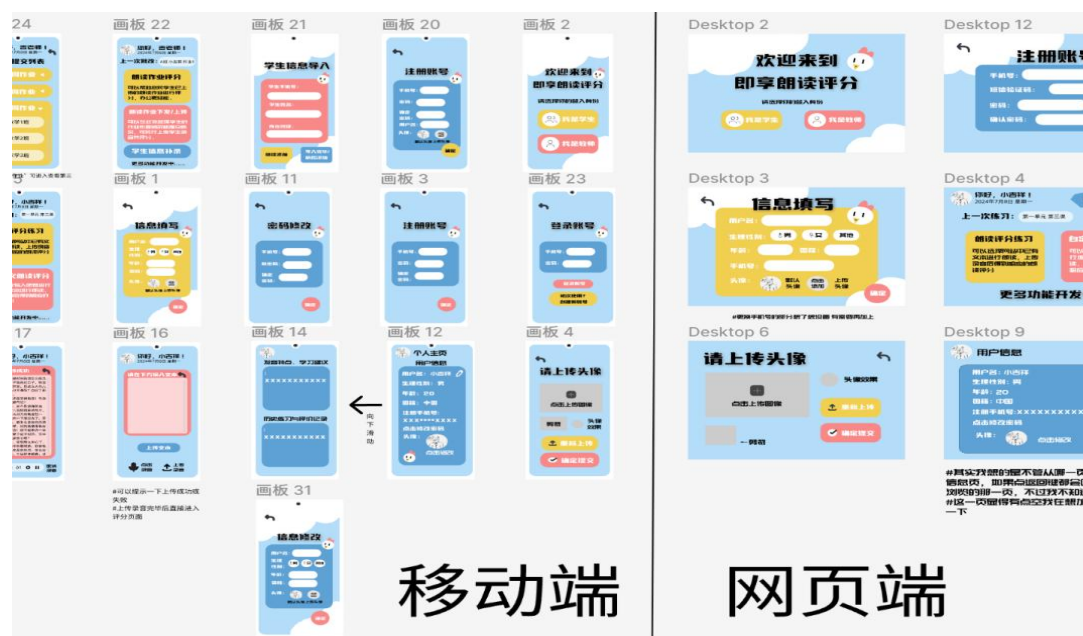


图 10 “即享朗读评分”移动端与网页端 UI 设计（部分）

目前，平台主要分为学生端（图 11）和教师端（图 12）两大板块，分别面向不同的用户群体。学生在填写基本个人信息后，可以进入学生端学习界面。该界面主要包括两大功能：一是朗读评分练习，即选择网站中已有的文本进行朗读并上传录音以获得相应的评分；二是自定义朗读评分，即自行输入文本进行朗读并上传录音以获得评分。在学生个人页面中，用户可以修改个人信息、进行个性化设置，并查看历史练习与评价记录，还可看到系统分析出的发音特点和个性化学习建议。教师端的设计则更为复杂，可用于布置朗读作业、辅助评分以及管理学生上传的录音等。鉴于开发时间有限，我们目前主要集中在学生功能界面的开

发上，未来有望根据已设计的教师端 UI 界面进一步完善其功能。



图 11 “即享朗读评分”移动端学生界面截图（部分）



图 12 “即享朗读评分”移动端教师界面截图（部分）

部分留学生在体验了该平台后对整体的用户界面、功能设计以及创新元素作出了积极的评价,并提出了一些建议,后续我们将依据这些建议作出针对性改进。

6、网页开发

目前我们运用多项技术,开发出了具备上述功能且可交互的网站(<https://chinese.redamancyxun.fun>)。网页开发细则如下。

(1) 技术栈运用

在本项目的开发过程中,我们采用了前端、后端及数据库等多层次的技术栈,以确保平台的稳定性、性能和用户体验(图 13、图 14、图 15)。通过合理选择和组合这些技术栈,我们构建了一个功能完善、用户友好的中文朗读评分与教学辅助平台。

1. 前端技术栈

前端部分主要使用传统的HTML、CSS和JavaScript三件套来构建用户界面,同时结合了一些现代的前端框架和库来提升开发效率和用户体验。

- **HTML5**: 用于构建页面结构,确保页面的语义化和可访问性。
- **CSS3**: 用于样式设计,实现页面的美观和响应式设计。
- **JavaScript (ES6+)**: 用于实现页面的动态交互和逻辑处理。
- **React.js / Vue.js / Angular**: 为了提升代码的可维护性和开发效率,项目前端已经开始框架化,未来将进一步完善构建框架。
- **Axios**: 用于处理API请求,提供简单的HTTP客户端功能。

- **Webpack / Vite**: 用于模块打包和项目构建,优化前端资源加载和构建过程。
- **ESLint / Prettier**: 代码规范工具,确保代码风格一致和质量。
- **Babel**: 用于JavaScript代码的转译,支持现代语法和浏览器兼容性。

图 13 项目开发使用到的前端技术栈说明

2. 后端技术栈

后端采用Spring Boot框架，提供了强大的Java开发支持，确保系统的稳定性和扩展性。

- **Spring Boot**: 基于Spring框架的快速开发工具，简化了Spring应用的初始搭建和开发过程。
- **Spring MVC**: 用于实现RESTful API接口，处理前端请求并返回数据。
- **Spring Data JPA**: 用于数据库访问，简化与数据库的交互，支持ORM（对象关系映射）。
- **Spring Security**: 用于用户认证和授权，确保系统的安全性。
- **Spring Boot Actuator**: 用于系统监控和管理，提供生产就绪的功能。
- **Spring Cloud**: 为了应对更多的用户量，项目后端已经开始微服务开发，未来将进一步完善构建微服务。
- **WebSocket / gRPC**: 用于实时通信，支持高并发的模型调用需求。
- **Git / DVC**: 用于代码和数据的版本管理，确保开发和生产环境的一致性。
- **JWT (JSON Web Token)**: 用于API身份验证，确保数据传输的安全性。
- **Logback**: 用于日志管理，提供系统的日志记录和分析功能。
- **Mockito / JUnit**: 用于单元测试和集成测试，确保代码质量和系统稳定性。

图 14 项目开发使用到的后端技术栈说明

3. 其他技术栈

除了上述主要技术栈，项目还涉及以下工具和库：

- **Docker**: 用于容器化部署，确保开发环境和生产环境的一致性。
- **Kubernetes**: 用于容器编排，提供高可用性和可扩展性。
- **Jenkins / GitLab CI**: 持续集成和持续交付工具，自动化构建、测试和部署流程。
- **Swagger / OpenAPI**: API文档工具，自动生成和维护API文档，方便前后端协同开发。

图 15 项目开发使用到的其他技术栈说明

(2) 网页核心功能和交互设计

目前网页的核心功能包括学生端的朗读评分练习、自定义朗读评分和个人主页，这些模块能够为学生提供一个高效、便捷的中文朗读学习环境，帮助他们提升口语能力，并通过详细的评分和改进建议提供清晰的反馈和持续改进的路径，实现有针对性的学习和进步。

首先，朗读评分练习是平台的核心功能，旨在帮助留学生提升中文口语能力。用户可以从卡片式布局的素材库中选择文本进行朗读，通过简洁的录音界面录制音频，并在上传后获得系统自动生成的评分报告。报告详细列出发音错误、语调

问题，并提供图文结合的改进建议。该功能提供高效便捷的学习体验，帮助用户通过有针对性的反馈实现口语进步。

其次，自定义朗读评分功能使学生可根据学习需求自行输入文本进行朗读练习，适用于练习特定词汇或句子。用户通过简洁的录音界面录制音频并上传，系统生成的评分报告提供发音和语调问题的详细反馈和改进建议。此功能满足个性化学习需求，进一步提升学习效果。

最后，个人主页模块集中展示学习记录和个性化设置。用户可通过时间轴布局查看每次练习的评分记录，包括练习次数、文本内容及建议。系统根据学习记录生成用户的“发音特点”和“学习建议”，以清晰易读的 Markdown 格式呈现。此外，用户可在主页修改头像、联系方式等个人信息，界面简洁，数据安全措施完善。此模块帮助用户了解学习进度，获取个性化改进建议，提升学习体验。

三、项目创新点

该项目在 **AI 大模型训练、语料处理、数据库建设、标注精细化、评分体系多维创新、以及用户界面设计**等六大环节体现了显著的创新点，具有较强的实践性和推广价值。

首先，本项目采用了**先进的 AI 技术**。我们应用智谱 AI 平台的 GLM-4-9B 模型，并采用 LoRA (Low-Rank Adaptation) 微调技术，对多项模型训练参数进行调整和优化，辅助以 Java 编写后端 API 和讯飞语音评测平台功能，初步实现了朗读语料评分和评语建议的智能生成。这使得机器在朗读语料评分合理度上以及语音评分效率上有所提升，同时节省了学生的学习成本和教师的人力劳动。

其次，在文本与音频数据的处理环节，本项目实现了**精准的数据切分和标注，并建立了高质量的语音评分数据库**。通过学院自建的韩国留学生语料库结合北大版《高级汉语听说教程》，项目团队实现了数据的多重细致分段与音频归一化处理，确保了数据的高效使用与精度提升。在音频分析中，利用讯飞非开源评分模型与 Python 接口进行自动化分析，并结合 Praat 进一步细化标注的创新性工作，弥补了声母和韵母具体偏误的标注空白。这一结合不仅提高了分析的精准度，还为未来大模型的细致评分提供了坚实的数据基础。此外，项目团队通过人工复检与格式化标注，如“/p/->/b/”和“[ɕ]->[ʃ]”，实现了系统的语内与语际偏误

区分，进一步推动了语音研究的细化和系统化。项目构建的数据库中包含不少于5000条可供模型学习的高质量音频，可以很好地应用于AI大模型的训练。

此外，在评分标准的制定方面，我们的**评分方式更加科学客观**。项目在机器评分的基础上，采用Z-Score和T值法实现了数据的量化和归一化处理，生成百分制评分，确保了评分结果的科学性和公平性。通过结合人工评分与机器评分的反馈，项目逐步完善了评分体系，创新地涵盖了准确度、流利度和完整度等多维度的分析，使评估更加全面和具有指导性。人工补充的声调标注尤其创新，将声调错误区分为“声调混读”和“怪调”类别，首次细致呈现了学生声调偏误的多样性。

最后，在用户体验方面，本项目设计的**平台简单易用、功能强大**。该项目设计了“即享朗读评分”平台，采用移动端优先的策略，充分考虑了用户碎片化时间的使用场景，使学生能够随时随地进行口语练习。此外，项目设计的吉祥物“小吉祥”所带来的人机互动设计提升了用户的学习参与度和使用体验。项目还综合了用户反馈并持续迭代优化，彰显了以用户为中心的设计理念。

综上，我们的项目在多个方面具有创新性，对既往的口语评分标准和市面上现有的朗读语料评分模型进行了完善与突破。

四、研究心得

本次创新项目的实施过程为团队带来了诸多宝贵的经验与启示，也在应对挑战与解决问题中留下了深刻的教训和反思。围绕自动化评分模型的构建，我们进行了包括文本分段、音频剪切与分析、评分标准制定、用户界面设计以及网页开发在内的多项核心任务。这些任务的成功实施不仅促成了项目成果的有效落地，也培养了我们的创新性思维，提高了我们的学习自驱能力、实践能力、团队合作能力和科研等能力。

在文本分段环节，我们深刻体会到合理的语段划分对于模型评分精确性的关键作用。通过实验验证，我们意识到文本的结构直接影响到朗读时的流畅度和自然性，从而影响评分的准确性。得益于此，我们在文本分段的过程中逐渐形成了一套系统化的方法，确保语段的逻辑性和易读性。该阶段的成功经验在于团队成员充分发挥各自的语言学和计算机技术专长，通过多次研讨与验证，最终确定了合理的分段策略，这为后续工作奠定了良好的基础。

音频剪切与分析的实施则是项目进展中最具挑战性的部分之一。在多次尝试与实验后，我们发现音频剪切的时长和方式直接影响模型的学习效果。过短的音频无法提供足够的语音特征，导致评分不稳定；而过长的音频又会增加计算负担，降低模型响应速度。经过细致的研究，我们为音频剪切制定了严格的规范，确保模型能够有效进行标准语音与学生朗读的对比。在此过程中，我们使用讯飞、praat 等语音分析工具，对音频的声学信号进行深度处理，提取发音的声母、韵母、声调、语速、停顿、完整度等多维度特征。该环节的技术挑战在于如何保证音频分析结果的稳定性和准确性，同时实现对语言学特征的精准捕捉。通过团队的共同努力和技术攻关，我们逐步克服了这些困难，提升了模型的评分效率和可靠性。

评分标准的制定是项目实施的重要节点之一。初期，我们尝试构建了综合评价体系，将机器评分与人工评分相结合，以确保评分的全面性和细致性。讯飞的评分结果提供了多维度的初始反馈，我们基于这些反馈信息进一步细化了评分算法，开发了涵盖不同发音特征的计算模块。然而，我们在此过程中也遇到了一些瓶颈，例如如何在人工评分中有效补充机器评分的不足，特别是在情感表达和语调细微变化的评估上。为了应对这些问题，我们在后期引入了逐项评估机制，使声母、韵母和声调的评分更加精确，并在情感维度上进行了初步尝试。虽然这一探索尚未完全落地，但它为今后的研究提供了有价值的思路。

UI 设计和网页开发是整个项目中将技术成果转化为用户友好体验的关键步骤。为了满足用户需求，我们设计了简洁、直观的界面，使学生能够轻松进行朗读测试并及时获得反馈。这一过程中，我们充分考虑了使用者的便捷性和功能的直观性，通过多轮的设计和用户测试，优化了界面布局和交互细节。此外，网页平台的开发为教师提供了便捷的教学辅助工具，使评分流程更加透明和高效，提升了教学效果和师生互动。

在更切实改进项目研究的过程中，我们细致了解了韩国留学生语伴们的需求和建议，这不仅使得两国学生的关系更密切了，还有助于国际中文教育实践的推广，与中国国际中文教育需求走在同一方向上。针对我们提供的朗读学练平台，大多数学生表示，该平台的移动端界面十分友好，易于操作，尤其是即时录音和上传功能极大地便利了他们在繁忙的学习和生活中利用零碎时间进行练习。一位

<p>学生提到，小吉祥作为平台吉祥物，不仅提升了学习的趣味性，还让他们感受到亲切和鼓励，这种寓教于乐的设计增强了他们对练习的兴趣。</p> <p>在项目实施的过程中，我们收获了许多值得借鉴的成功经验，例如团队间的协作和跨领域知识的融合。但同时，我们也认识到一些失败教训，如在早期阶段对音频分析的预估不足，导致反复调整 and 资源的浪费，以及评分标准在情感维度上的不足，需要在今后的研究中进一步完善。整体而言，这次项目不仅提升了我们的创新能力和技术水平，锻炼了我们的科研思维，还在如何系统化、全面地实施一个复杂项目方面提供了宝贵的实战经验。</p>			
五、项目组成员			
姓名	学号	专业	项目研究中承担的主要任务
于佩民	10215102460	汉语国际教育	语音评测和标注算法的协助开发，负责优化和提升基于讯飞语音和面向计算机辅助正音的汉语中介语语音语料库的标准化标注系统。
黄奕程	10222740415	汉语国际教育	负责优化朗读评测模型，确保其在特定语料下的准确性和可靠性，根据实际数据对模型进行调整和补正。
陈松涛	10222740450	汉语国际教育	分析语音声学数据，以便为教学提供有效的参考和指导，权衡标注参数和调配权重，以确保评测结果符合教学需求。
叶晓良	10235101427	软件工程	负责设计整个系统的前后端架构，包括数据库调用、模型设计与算法开发、接口开发等操作系统的功能实现方面的工作。

六、经费使用情况及下一步研究计划

1、经费使用情况

目前使用经费情况如下情况如下：

阿里云服务器（1 年），用于平台网站运行

百度网盘 SVIP 会员（1 年），用于保存语料库音频和文本

GLM-4-9B 模型，用于项目模型训练

2、下一步研究计划

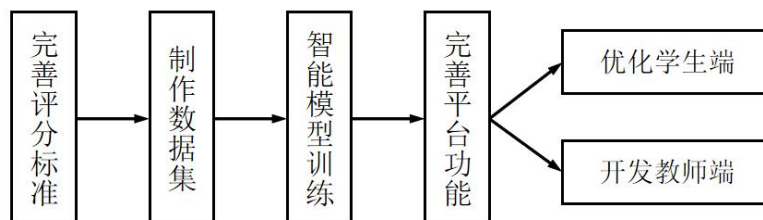


图 16 技术开发路线图

后续研究工作中我们计划从以下几个方面进一步深入探索：

（1）完善评分标准

我们现阶段的评分标准在实际应用中仍存在一些针对性问题和不合理之处，涉及评分者和学生两方面。对于评分者而言，过于细致的评分标准在实际中并不实用。现实环境下，对学生朗读准确度的评价通常不会细化到声母、韵母和声调层面，而是以音节作为更易识别和判断的单位。因此，从整体上对声韵调进行评分更为合理。此外，评分者通常基于整体印象进行评分，各维度间会相互影响。例如，当声韵调表现较差时，即使流利度良好，评分者也倾向于给低分，认为学生为求速度而牺牲准确性。整体印象还会受到朗读者情感、音色等边际因素的影响。基于这些考量，未来的评分设计应在不影响模型训练的前提下，平衡模型需求与实际应用，使其更符合人类认知并便于操作。

从学生角度来看，成绩在 0 到 100 范围内并非正态分布。通常情况下，当学

生在班级或集体中时，评分依据的是学生间的分布而非绝对正确率。考虑到留学生的学习环境，确保大部分学生取得较高成绩至关重要。基于此，我们制定了调整方案。

当前评分结果已被分类（见图 17），其中 76.25-100 和 55-76.25 分区间的音频各占 31%，合计 62%。这意味着 62% 的学生能取得百分制 80 分以上或五级制 4 分以上的成绩，这被认为是合理的分布。因此，我们将 76.25-100 区间对应 90-100 分，55-76.25 对应 80-90 分，36.25-55 对应 70-80 分，20-36.25 对应 60-70 分，0-20 对应 44-60 分，最低分数限定为 44。

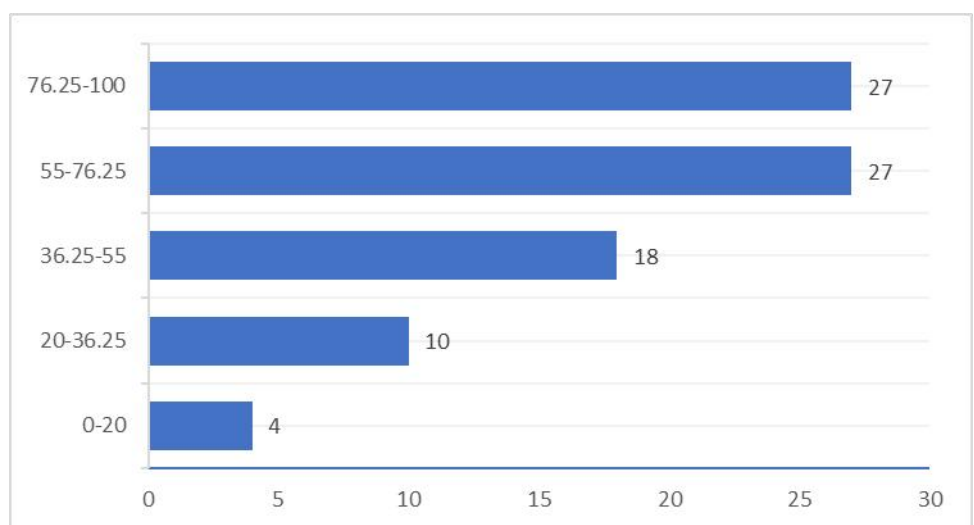


图 17 学生朗读语料成绩分类统计图

基于这一对应关系，我们将分数计算的方法调整为下面的公式，并期待在下一阶段进行检验。

$$Result = \sqrt{\left(\frac{X - MIN}{MAX - MIN} \times 80 + 20\right) \times 10}$$

（2）研究论文或报告撰写

经过现阶段关于韩国留学生语料的研究分析，结合大模型对人工评分的学习，我们已能初步总结出影响韩国学生朗读语音评分的部分因素，以及这些发音中的常见偏误对机器朗读评分的影响，也一定程度上量化了机器评分相对人工评分更为“生硬”的来源。在后期研究中，我们计划采集更多韩国乃至其他国别留

学生语料，依托模型训练和 AI 数据反馈，进一步分析造成外国人语言面貌不够地道的主要因素，并从朗读评分的角度展开，一方面对学生提出可尝试的提升路径，另一方面对现有的机器评分规则提出可改良的评分算法，有望出产相关论文或报告一篇。这不仅有助于提升评分模型的准确性，还可以对汉语学习的偏误分析理论有所贡献。

（3）拓展平台功能

①优化学生端

为更好服务学生利用软件完成自学自纠和得到切实进步，我们结合国际中文学生群体的建议，考虑在提供给学生能够反映当前绝对水平的百分制或五级制分数的基础上，对产品做出如下改进：

一是要**继续丰富平台提供的中文朗读材料**，以涵盖不同的难度和主题，保证朗读文章选材的权威性、时效性、趣味性、热点性和生活性，使学生能够根据自己的中文水平和学习目标进行更有针对性的练习。我们计划选择《发展汉语》作为后期语料来源。该教材为对外汉语长期进修教材，普通高等教育“十一五”国家级规划教材，采取“综合语言能力培养与专项语言技能训练相结合”的外语教学及教材编写模式，分为“三个层级、五个系列”，即纵向分为初、中、高三个层次，横向分为综合、听、说、读、写五个系列。其中关乎到本项目的语音、口语教学方面，选材精当合理，作为朗读材料，较能够体现出学生的朗读水平。更重要的是，该教材为我校外国留学生汉语学习的主要教材，可以更加方便我们获得语料。

二是**加入激励性功能**。语言学习重在重复和坚持，如何避免“三分钟热度”，从而让学生真正带着热情投入语言的学习是需要不断思考的命题。我们希望以图表形式呈现学生的学习效果，如进步曲线图或得分变化趋势图，从而让学生更直观地看到学习成效，增强其自信心和学习动力。还可以引入每日签到打卡、积分兑换礼品和经验等级模式，使学习者有机会在产品内部找到某一奋斗目标，从而将其转化为语言学习的动力，进而实现产品使用频率和语言水平的提高。

三是面向学习者**提供有针对性的标准音频和舌位动画**。对目前的国际中文语

音教学而言，让学生直接模仿正确读音是最高效经济的纠音方法。如果只反馈给学生分数，除了减轻老师的评改负担外，学生并不能利用该系统获得很好的指导性改进意见，不利于学习者进一步提高。因此，通过分析学习者所提交的朗读音频生成的量化信息，该产品将能为用户提供其典型错误的正确读音示范，从而帮助学习者自查自纠。对于那些难发难纠的音，还应适时提供舌位动画，便于学习者进行理解。

四是要引入**相对性评价**。通过设计“每日阅读比拼”模块，吸引众多不同等级的学习者同读一篇文章，并设置实时动态榜单，允许学生多次尝试朗读该文章，帮助学生更好地确定自己在学习者当中的相对水平，通过趣味竞争激发学生学习的积极性和主动性。

五是要利用众多的用户语音数据**不断调试和改进现有模型**。大模型的不断精进建立在高质量、大规模的训练集基础上，基于用户使用产品而获得的数据具有真实性和普遍性，有助于进一步优化现有模型对于评分的准确性和更好实现其对部分主观评分项目与语音数据关联性之间的自动分析。

②构建教师端

在完善学生端构建的基础上，我们期望对教师端进行相应的开发。教师端的设计与实现相较于已有的学生端而言具有更高的复杂性和多样性，其核心功能包括朗读作业的布置与检查提醒、辅助评分以及学生上传录音的管理等，我们希望构建的方面如下：

一是**完善朗读作业的布置与管理功能**。教师可以设定朗读材料、明确任务要求，并指定完成的时间节点。学生完成朗读任务后，教师可以通过系统方便地查看学生提交的录音文件，及时给予反馈。

二是**构建辅助评分功能板块**，通过自动化评分模型提供初步的评分结果，减少教师的评分负担。教师可以对评分结果进行进一步的调整和细化，尤其是对于难以通过自动化评分捕捉的细节（如语调的自然度、语音流畅性等）进行手动评分。这一功能将大大提高评分的准确性与公平性，确保评分过程既具备高效性，又不失个性化与专业性。

三是**加入录音文件的批量管理与分析功能**，使得教师能够方便地浏览学生提交的所有朗读录音，并对录音进行分类、标签化管理，以便后续的复查与教学规划。此外，系统应支持对录音文件的快速检索与筛选，帮助教师在众多学生录音中迅速找到需要重点关注的部分，从而提高教学针对性和精准度。

四是**增加大模型支持的个性化教学工具**。随着数据量的不断积累，教师端可以通过接入 ChatGPT 等大模型，对学生的学习进度与口语表现进行更深入的分析，为教师提供更加个性化的教学建议。例如，系统能够自动识别出某些学生在特定语音问题（如发音准确性、声调变化等）上的持续困难，并向教师提供具体的干预建议，帮助教师根据学生的需求调整教学内容和方式。

（4）计划为本项目中建立的“即享朗读评分”平台申请软件著作权。保护软件的原创性设计。

（5）用户测试计划

首先，在中期报告前后，完成产品自动评分功能的基本实现，同时利用项目资金，借助学院丰富的留学生资源，开展产品内测，进入我院语言课堂（语言班1-4D），在不干涉教师正常教学进度的情况下推广平台，丰富数据并结合内测反馈及时改进和完善现有模型。

其次，广泛听取专家意见。计划邀请经验丰富并长期奋战在一线教学岗位的国际中文教师，特别是听说课或口语课教师参与产品性能分析，尤其是要关注评分结果的合理性、评语的有效性和偏误分析的准确性；要确保评分和反馈能够切实指导学生提升。

最后，在初步完善的基础上，扩大测试规模。在多平台上对“即享朗读评分”平台进行推广，邀请多学段、多层次的汉语学习者进行产品体验，有效推进产品普适性的提升，争取和上海各大高校相关专业建立合作关系。

七、指导教师意见（请对项目进展情况作出评价，并对下一步研究计划提出建议）

我认为该项目在多个方面均表现出色，值得充分肯定。

在项目工作量方面，团队成员们展现了极高的工作热情和责任心。无论是数据标注的繁琐工作，还是代码编写的技术挑战，都得到了妥善处理。这实属不易。在技术创新和功能模块设计方面，项目团队也展现出了令人赞赏的创意和实力。他们不仅成功构建了评分模型，还在此基础上进行了多次优化，使得软件的功能更加完善，用户体验更加友好。特别是在界面设计上，兼顾了简洁性和易用性，使得软件在初步实用阶段就得到了较好的反馈。

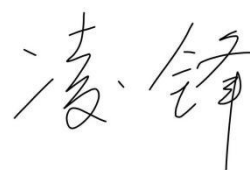
总之，能够在如此短的时间内完成一个基本功能齐备、已经可以初步应用的软件，体现了团队成员们的专业素养和团队协作能力。

当然，项目仍有进一步提升的空间。我建议团队在接下来的工作中，进一步优化评分体系，提高评分的准确性和稳定性。同时，可以邀请更多使用者进行试用，收集他们的反馈意见，以便更好地完善软件的功能和用户体验。

此外，结合大模型对人工评分的学习，团队还可以进一步分析影响外国人语言面貌不够地道的主要因素。这不仅有助于提升评分模型的准确性，还可以对汉语学习的偏误分析理论有所贡献，为未来的汉语教学和研究提供新的思路和方法。

综上所述，“国际中文朗读语料评分模型建构与优化实践”项目表现出了较高的水平和潜力。我相信，在团队成员们的共同努力下，项目能够顺利进行下去。期待他们的后续工作。

指导教师（签字）：



2024 年 11 月 11 日