

一款汉语智能语音教学 App 及教学实验初步结果 (An Intelligent Chinese Pronunciation Teaching App and the Preliminary Result of a Teaching Experiment)

魏巍

(Wei, Wei)

北京语言大学

(Beijing Language and Culture University)

wwei906@163.com

张劲松

(Zhang, Jinsong)

北京语言大学

(Beijing Language and Culture University)

zjsblcu@qq.com

摘要: 发音教学一直是汉语作为第二语言教学中的难点问题。近年来迅速发展的人工智能发音教学技术 (CAPT) 能够自动诊断学生发音中的偏误, 并给出正音的引导信息, 从而帮助学生不受时空限制的完成自主发音学习, 被认为是克服二语发音教学问题的希望技术。但是由于出现时间尚短, 技术也有待完善, 其教学辅助效果到底如何是一个令语学界感兴趣的问题。北京语言大学智能语音习得技术实验室一直以来从事汉语作为第二语言的智能语音教学技术的研究, 开发出以尔雅中文教材为模板的汉语 CAPT-APP 软件。为了探索 CAPT 实效问题, 本文以该软件为工具, 在北京语言大学的两个留学生班级进行了为期 6 周的教学实验。初步的实验结果显示, 36 名学生在 6 周的时间内使用 APP 练习发音的总次数为 28101 次, 平均每名学生练习次数为 780.6 次, 其中, 课后练习次数占到了 91.9%; 发音偏误细节反馈信息能够帮助学生改善 83.2% 的发音问题, 并且改善一个发音问题平均只需要约 1.5 次的重复练习。这样的结果说明 CAPT 教学工具对于二语语音教学的确非常有效, 不仅能够有效辅助学习者完成大量发音练习, 而且有助于他们改善发音质量。

Abstract: Pronunciation has always been a challenge in the learning and teaching of Chinese as a second language. Two common best practices to improve learners' pronunciation include providing learners with sufficient meaningful practice opportunities and quick feedback from their instructors. However, due to limited classroom time and one-to-many relationship between teachers and students, it is difficult to follow the above two practices in classroom instruction. Driven by artificial intelligence technology, intelligent computer assisted pronunciation teaching (CAPT) system can help learners to practice pronunciation on their own and receive instant feedback. This paper reports the findings of a six-week teaching experiment in using *Erya* Chinese pronunciation APP at Beijing Language and Culture University. Our preliminary results showed that the total number of pronunciation exercises was 28,101 times, with an average of 780.6 times per student, of which the number of after-school exercises accounted for 91.9% in quantity; under the guidance of detailed feedback

information, students can remedy 83.2% of their pronunciation errors efficiently by only 1.5 exercises on average. This result showed that the intelligent CAPT system can help the learner to complete a large number of pronunciation exercises and improve the pronunciation quality.

关键词: 二语教学, 汉语发音, 智能教学技术, 教学实验

Key words: Second language teaching, Chinese pronunciation, intelligent teaching technology, teaching experiment

1. 引言

语音教学是二语教学的首要任务, 是二语教学活动开始的第一个环节, 也是一个重要环节。顺利习得正确的二语语音系统有助于激发学生对于后续学习的兴趣, 构造起语言能力和社交能力发展的坚实基础。而发音的偏误则很容易被母语者感知到, 不仅会带来强烈的外语口音感, 还可能影响到言语的可懂度(张林军, 2015)。因此, 在针对母语非汉语的汉语教学活动中, 师生都需要重视汉语基本语音系统的教授和学习。

语音教学也是个难点问题, 其表现就是在二语学习者发音中普遍存在的非母语腔调问题。在对外汉语教学中, 则被周知为“洋腔洋调”现象。不论是汉语的声母、韵母, 还是声调等基本语音要素, 对于母语非汉语的学习者来说都存在着各种不正确的发音现象, 而且这种发音偏误在不同国家的学习者中又有着不同的表现(王韞佳等人, 2004; 吴门吉等人, 2004; 梅丽, 2005; 王茂林, 2011)。

关于二语语音习得困难的原因, 学界已经有很多的研究。其中代表性的理论——语言习得关键期假说(Lenneberg, 1967)认为, 人类的语言习得过程存在着最佳时期, 在这段时期内, 大脑认知机制最有利于语言习得和发展, 一旦过了这段时间, 认知机制就会发生变化, 后续的语言学习就会变得困难。近期研究进一步从脑神经的工作过程揭示了不同年龄学习者存在脑部语言加工位置不同的事实(Kim 等人, 1997)。由于人脑认知、加工机制带来了成人语音习得的困难, 所以, 大量的练习对于成年学习者来说具有必要性。

对于成年人来说, 虽说语音学习比较困难, 但是成人经过训练仍然可以在二语语音的知觉和产出上取得进步(梅丽, 2014)。二语习得理论的输入假说理论便认为(Krashen, 1985), 学习者在接收“可理解性的语言输入”后, 便可能触发二语能力的习得。针对大脑认知机制发展的研究也表明, 即使过了儿童二语学习的敏感期, 大脑仍然具有一定的可塑性, 可以通过外在刺激的使用引发语言能力的发展(Li 等人, 2014)。相关的研究和实践表明, 即使过了二语学习的最佳时期, 学习者也可以通过大量的练习来改善学习效果、获得二语的语音能力(张林军, 2009; 孙悦等人, 2013; Feng 等人, 2014; 李艳萍等人, 2017)。

针对如何更加有效地获得二语语音能力的问题, 输入假说理论提出“i+1”语言输入原则 (Krashen, 1985): 提供给学习者的语言材料应该是循序渐进的, 是学习者容易理解和掌握的, 即“i”部分 (“i”代表 *intelligible*), 并在已经习得内容上增添部分新的内容, 即“1”部分, 代表一部分新的内容。在语音能力训练过程中, 关于输入语言的意义或者是偏误原因的反馈信息, 都有助于学习者掌握所要学习的内容, 有利于学习者从一个阶段进入另一个更高的阶段。

这样的理论表明, 语音教学活动需要满足两个条件: 一是学习者需要进行充足的有意义的发音练习; 二是在练习时学习者能够及时收到有针对性的正音反馈信息。例如在汉语发音教学中, 关于 VOT 特征 (*Voice Onset Time*) 的及时反馈有助于学习者改善送气、不送气辅音的混淆问题 (张林军, 2009), 关于音高调域高低的反馈有助于解决阳平、上声的混淆问题 (孙悦等人, 2013; Feng 等人, 2014), 关于第二共振峰变化的反馈有助于解决前后鼻音的混淆问题 (李艳萍等人, 2017), 等等。这些运用音频编辑技术的语音训练方法, 通过改变关键声学线索, 帮助学习者有针对性地对易混淆的两个音, 并通过足量的练习改善发音问题。

在传统的课堂教学活动中, 语音教学是师生间通过语音互动活动开展的。由于师生数量比例往往是一对多的关系, 语音教学便只能通过顺序、分时进行。加上课堂时间总量有限, 每位学生分摊到的训练时间有限, 便难以得到充分的发音机会, 也难以得到来自教师的及时且充分的正音反馈意见。因此, 在课堂教学实践中, 语音习得的两大必备条件都难以满足: 学生难以在教师指导下进行大量的发音练习, 也难以获得及时的正音反馈指导。这样一个由教师资源稀缺所造成的语音教学困境, 进一步阻碍了二语语音教学的效果, 加剧了二语语音习得的不正确腔调问题。

近年来迅速发展的计算机辅助发音教学 (*Computer Assisted Pronunciation Teaching, CAPT*) 日渐引起人们的关注 (Godwin-Jones, 2011; Yang, 2013; Liu 等人, 2016)。首先, CAPT 系统能提供丰富的多媒体资源, 如标准语语音的音频、发音动作讲解视频等信息, 有助于增加学习者的二语语言资料输入量。其次, 智能化的 CAPT 系统还能够评估学习者的发音质量, 有助于学习者自主进行大量的发音练习。例如基于网络多媒体技术的“长城汉语”系统¹, 基于移动互联网的微软中文学习 AI 助手 *Learn Chinese*²等。一项问卷调查的结果显示 (刘洋, 2018), 在被调查的本科及以上学历的外国留学生和孔子学院的学生中, 有超过 90% 的学生都下载过 APP 汉语辅助教学软件学习中文, 学生对语音问题有较高的倾向性, 多数学生认为汉语辅助教学软件可以帮助他们解决语音问题。

但是, 由于智能 CAPT 技术出现时间尚短, 其技术特征本身尚存在很大的待提升空间 (Levis, 2007)。目前绝大多数面世的智能 CAPT 技术所提供的反馈信息以发音质量评价信息为主, 具体形式以整句评分或者表征发音质量“好/坏”的不同颜色形式为主。这样的反馈信息虽然能够帮助学习者了解发音是否存在问题, 但是由于缺少关于具体发音问题的细节反馈, 学习者并不知道如何去修正发音, 从而影响

¹ <http://www.greatwallchinese.com.cn/portal.do?method=user>

² <https://www.msra.cn/zh-cn/news/features/microsoft-learn-chinese-20171127>

了发音学习效果。另外, 由于缺少较为严格的教学实验验证, 人们对于智能 CAPT 技术的教学效果也心存疑问, 不知道作用到底如何。

本论文旨在通过教学实验来研究上述关于智能 CAPT 技术教学效果的问题。所采用的“尔雅中文”汉语智能语音教学 APP 由北京语言大学智能语音习得技术实验室 (SAIT Lab) 研制, 在北京语言大学初级留学生课堂开展了为期 6 周的使用实践。该款 APP 利用大规模的汉语中介语语音语料库 (曹文等, 2009), 运用深度神经网络模型 (Deep Neural Network, DNN), 能够实现较为精准的具体发音偏误检测, 不仅能够就整句发音进行质量评分, 还能够就具体的声母、韵母及声调偏误进行检测 (Gao 等人, 2015; 张劲松等人, 2016; Wei 等人, 2017; 屈乐园等人, 2017; Lin 等人, 2018), 并以颜色提醒、声音对比、动画展示、文字描述等多种方式给予发音教学反馈。

以下内容安排如下: 第 2 节介绍“尔雅中文”APP 所实现的主要功能, 第 3 节介绍发音教学实验设计, 第 4 节介绍实验结果, 第 5 节给予简单结论。

2. 智能语音教学 APP——尔雅中文 APP

从满足发音教学需要出发, 尔雅中文 APP 设计了学生端³和教师端两大模块, 各模块主要功能如图 1 所示。

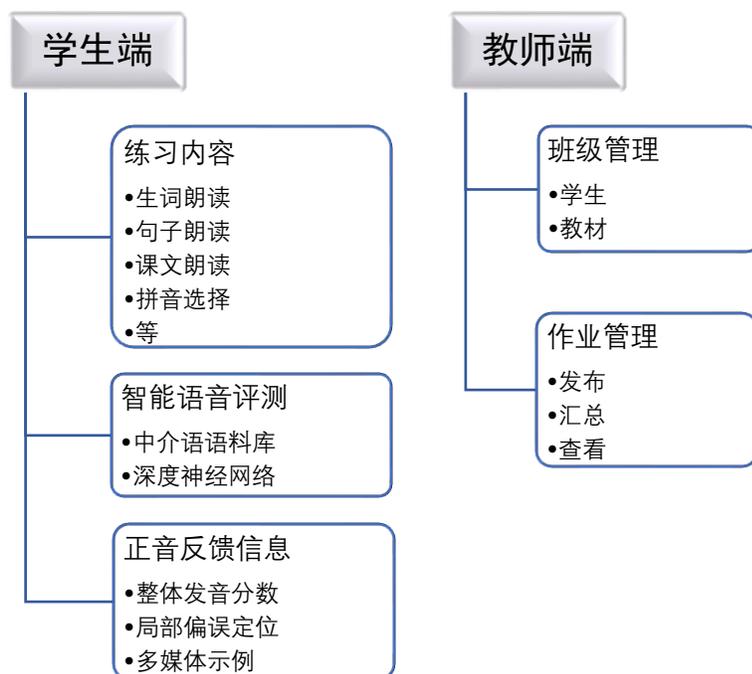


图 1. 尔雅中文 APP 功能模块结构图

³ 尔雅中文 APP 学生端安装方式: 1. 安卓版可在 <https://fir.im/tyzd> 下载; 2. 苹果版可在 App Store 搜索“SAIT 汉语”获取。

首先, 学生端收录了汉语教材《尔雅中文》⁴中的练习材料, 给学生提供生词朗读、句子朗读、课文朗读和拼音选择等练习, 如图 2 (a) (b) (c) 所示。以发音练习模块为例, 练习的初始页面呈现练习内容, 并提供示例音频。学习者可以参考文字提示和音频示例, 在手机上录制发音练习的音频, 然后就会收到 APP 对发音进行评测后的反馈信息。



图 2. APP 模块示例图

其次, 智能语音评测模块依托汉语中介语语料库, 运用基于深度神经网络的语音评测技术, 对学习者的发音进行整体打分和局部偏误的定位检测。发音评测的基本依据是学习者的发音和标准语音模型在声学上的相似度, 如果相似度较低, 则判断为存在偏误。APP 会识别学习者练习发音的音频内容中每个音节的声、韵、调, 并对每个音节的声、韵、调进行相似度打分, 依此对每个声、韵、调的正误进行判断, 并得到整体发音的分数。

⁴ 魏新红. (2013). *尔雅中文: 初级汉语综合教程 (上、下册)*. 北京: 北京语言大学出版社.



图 3. 尔雅中文 APP 应用示例

再次, APP 采用“整体评分+局部偏误”的发音信息反馈方式, 引导学习者纠正发音。以一名女性韩国学生的发音练习数据为例, 如图 3 (a) 所示, 学习者所发的“旅行社”的第一个音节的声母“l”存在发音偏误, APP 用红色区分有问题的发音, 点击有偏误的声母, 会收到图 3 (b) 所示的反馈信息, 向学习者说明此处将“l”的发音混淆成了“r”。同时, APP 会向学习者反馈发音的整体分数。当学习者收到发音偏误的反馈信息后, 可以有针对性地练习存在偏误的声、韵、调。如图 3 (a) 和图 3 (b) 所示, APP 提供了标准语音的音频, 发音动作演示视频等多媒体教学资源, 学习者可以通过反复听辨、视觉对比、发音模仿等方法来对比正误发音, 通过自主式的反复摸索练习, 进而掌握正确的发音方法, 改善发音质量。如图 4 所示, 图 4 (a) 是从图 3“旅行社”的语音中提取的“旅”字的语图, 图 4 (b) 是学习者纠音后“旅”字的语图, 图 4 (c) 是从中介语语料库中选取的女性中国发音人单音节“旅”字的语图。感知实验研究表明 (陈静萍, 2018), 过渡段对中国被试感知声母 l、r 具有关键作用, 普通话 l、r 的声学特征主要差异表现为过渡段的第三共振峰 (F3) 的模式, 由于卷舌的作用, r 的过渡段的 F3 具有较大的变化范围。从图 4 中可以看出, 学习者通过纠音练习, F3 的模式向母语者的模式靠近, 这反映出了学习者发音质量的改善。

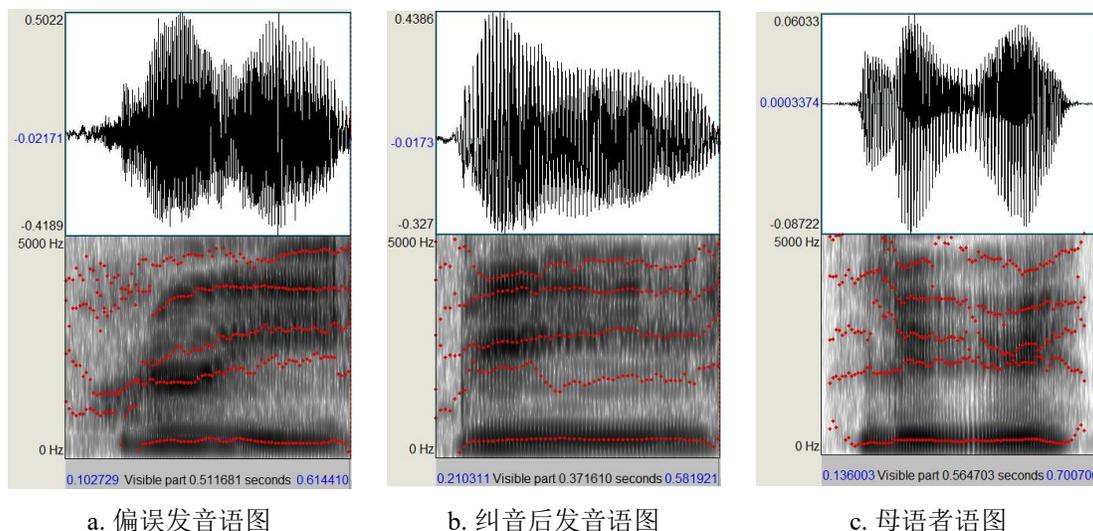


图 4. “旅”字声母纠音前后对比图

同时, APP 还有专门的教师端, 以班级群组的形式管理课堂教学。教师在新建班级后, 可以选择要使用的教材。目前, APP 收录了《尔雅中文: 初级汉语综合教程》上、下两册, 后期会根据教学的需求扩充教材资源。学生通过申请加入班级后, 就可以使用指定教材的内容来练习发音。教师可以通过 APP 向班级中的学生发布作业, 如图 2 (d) 所示。学生提交作业后, 教师可以查看学生的作业完成进度, 以及收听学生练习发音的音频。

3. 发音教学实验设计

3.1 实验目的

本研究采用行为数据分析的方法, 考察二语学习者使用尔雅中文 APP 练习发音的情况, 以探究智能语音教学工具在二语教学中对发音教学的作用。

3.2 实验原理及评价指标

SAIT Lab 研制的基于 DNN 的智能语音评测技术, 在实验的环境下, 检测准确率平均约为 90.4% (张劲松等人, 2016; 屈乐园等人, 2017; Lin J.等人, 2018)。APP 的智能语音评测模块就是在这些已有的技术积累下研制而成的, 可以根据声、韵、调的相似度来检测发音偏误。发音练习是语音习得的一个过程, 在这个过程中, 发音质量是在一个范围内波动变化的。智能语音评测技术可以通过发音的分数将这个波动变化的过程进行量化。发音的分数反映了发音与标准语音的声学相似度, 当相似度较高, 即认为发音接近母语者的发音; 当相似度较低, 即认为发音存在明显偏误。学习者发音的评测结果从“有偏误”到“无偏误”, 能够客观反映学习者的发音质量得到了较明显的改善。学习者根据 APP 的发音正误反馈进行的纠音练习, 就是通过练习不断地改善偏误发音, 最终接近标准语音的过程。

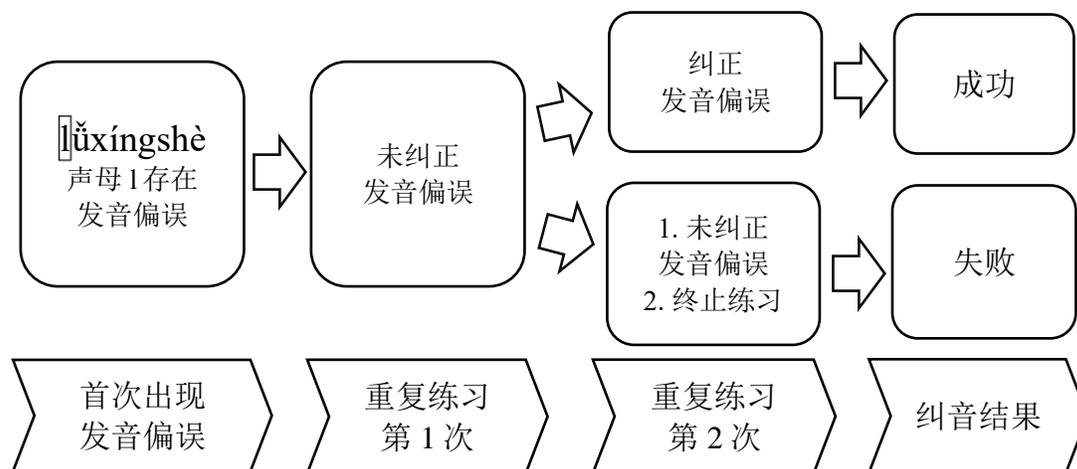


图 5. 纠音练习流程示意图

本实验考察学生使用 APP 的发音练习量和纠音效果。学生使用 APP 练习发音时, 每练习一个条目, 就会录制一条发音练习的音频文件, 因此, 音频文件的总数即是学生练习发音的总次数, 音频文件的总时长可用于衡量发音的总时长。发音练习量以练习的次数和时长进行统计。纠音效果主要是为了反映“局部偏误”的反馈信息对学生纠正发音的有效性, 以及学生的练习效果。

在练习发音的过程中, 学生会根据 APP 的反馈信息对有偏误的发音反复地练习, 以纠正发音偏误。实验中, 我们先从发音练习的音频文件中提取学生重复练习的词条; 再从这些重复练习的词条中, 整理出每一个音节的声、韵、调的重复练习过程, 并从中提取学生的纠音行为。如果通过反复练习将偏误的音改正, 则视为纠音成功; 如果最终没有改正, 则视为纠音失败, 如图 5 所示为“旅行社”这个词条中声母“l”的纠音过程。从首次出现发音偏误到结束纠音的过程称为一个纠音练习过程, 记作 E , 其中, 纠音成功的记作 E_T , 纠音失败的记作 E_F ; 在 E 中, 重复练习的次数称为纠音练习次数, 记作 N , 图 5 所示纠音过程的练习次数 $N=2$ 。

目前, 多数 CAPT 系统关注的是技术的创新, 而不是教学的标准 (Neri 等人, 2002)。我们从改善发音质量的实际教学目标出发, 针对纠音练习的具体过程, 采用纠音成功率和纠音成功练习次数作为评价纠音效果的指标。由于纠音过程是在反馈信息的提示下进行的, 因此, 这两个指标能够直接反映反馈信息的作用。纠音成功率 P 从整体上反映所有 E 中 E_T 所占的比例, 定义如式 (1) 所示:

$$P = \frac{T}{T+F} \times 100\% \quad (1)$$

其中, T 为 E_T 的总个数, F 为 E_F 的总个数。纠音成功练习次数反映的是 E_T 的具体细节, 即纠正一个发音偏误所需要的重复练习次数, 记作 N_T 。实验将从 P 和 N_T 两个指标分别评价声母、韵母和声调的纠音效果。

3.3 被试及实验语料

参加本次发音教学实验的是北京语言大学两个班级的留学生, 共 36 名, 分别

来自 19 个国家, 各国家学生人数如表 1 所示, 学生的汉语水平为初级。

表 1: 各国学生人数统计表

国家 (同一格内按音序排列)	人数 (单位: 人/国家)	小计 (单位: 人)
韩国	7	7
巴基斯坦、俄罗斯	4	8
泰国、意大利	3	6
澳大利亚、乌兹别克斯坦	2	4
丹麦、哈萨克斯坦、墨西哥、斯里兰卡、 特立尼达和多巴哥、委内瑞拉、乌克兰、 西班牙、伊朗、印度尼西亚、英国	1	11
总计		36

APP 自动记录了学生练习发音的相关数据, 其中包括学生练习发音的音频文件, 音频对应的发音练习文本(汉字及拼音), 机器对音频语音识别的结果, 机器对学习者发的每一个音节声、韵、调的打分及正误判断, 机器对整体发音质量的打分, 以及练习的时间等信息。从这些数据中可以得到学生使用 APP 练习发音的行为数据, 如使用时间, 纠音过程等。

通过对采集到的音频进行抽样调查, 发现整体发音成绩小于 40 分的音频大多是由于录音操作不规范导致的音频未录完整。在分析发音质量时, 将自动剔除得分小于 40 分的音频, 这部分音频约占数据总量的 3.2%。

3.4 实验过程

实验在汉语课语音教学周期内同步进行, 两个班的汉语课都安排在星期一至星期四, 每天两课时, 为期约为 6 周。在教学实验周期内, 学生通过 APP 自主地完成发音练习。教师会根据授课进度和教学需求, 通过教师端发布作业, 并指定提交作业的截止时间, 引导学生完成指定的发音练习任务。整个教学周期内, 所有发音练习的行为数据均由 APP 自动地采集和记录, 学生作业的成绩将计入最终的学业考核成绩中。

4. 教学实验初步结果和讨论

4.1 发音练习量统计结果

图 6 表示的是在整个教学实验周期内, 所有学生使用 APP 练习发音的总次数, 以及课堂内外使用 APP 所占的比例。图 6 (a) 反映了所有学生累计使用 APP 练习发音的总次数为 28101 次, 人均练习约 780.6 次。所有发音练习的音频总时长为 21.0 小时, 平均每个练习的音频约 2.7 秒。图 6 (b) 说明课堂外的练习次数占总练习次数的 91.9%, 课堂内的练习次数占总练习次数的 8.1%。

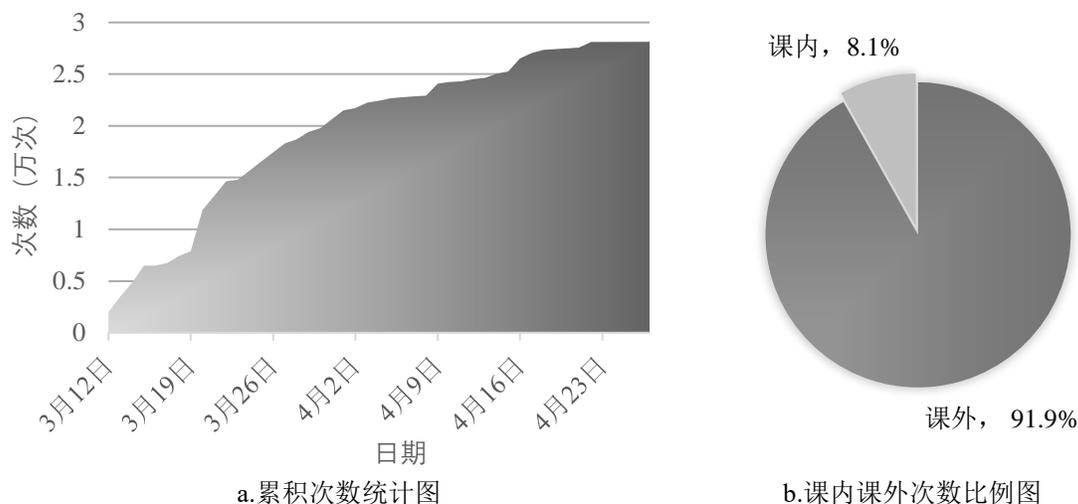


图 6. 发音练习次数统计图

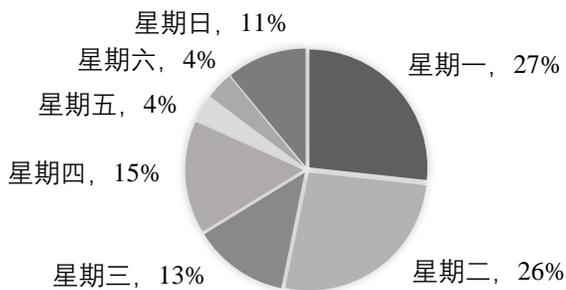


图 7. 发音练习次数按星期分布比例图

图 7 反映了发音练习次数按星期的分布情况, 周一、周二、周三、周四和周日的练习量较多。教师端记录的数据显示, 两个班的教师共发布了 31 次作业, 发布时间为授课当天, 其中有 20 次作业是要求学生当天完成, 有 9 次作业学生需要在 1 至 2 天内完成, 还有 2 次周四布置的作业要求在周日前 (包括周日当天) 提交。学生收到作业后, 习题的完成量平均约为 88.7%。

4.2 纠音效果分析结果

行为数据中提取到的纠音练习过程 E 的总数约为 4.3 万个, 其中, 声母、韵母和声调的比例分别如图 8 所示。

声母、韵母和声调的纠音成功率 P 分别如图 9 所示。从图中可以看出, 声母和韵母的 P 相近, 均高于 84%; 声调的 P 相对偏低, 略低于 80%。

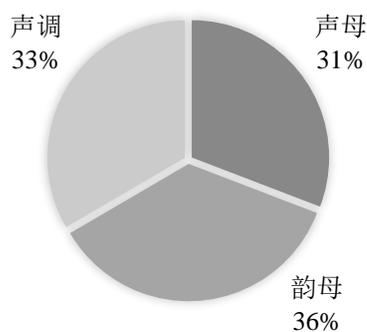


图 8. 纠音练习总个数比例图

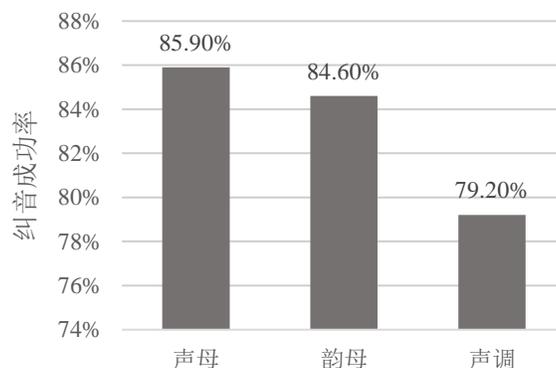


图 9. 纠音成功率统计图

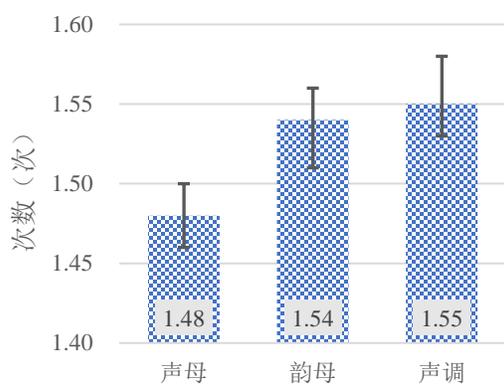


图 10. 声、韵、调纠音成功练习次数统计图

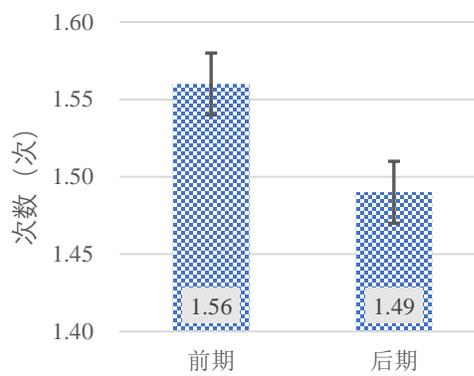


图 11. 前、后期纠音成功练习次数统计图

图 10 表示的是声母、韵母和声调的纠音成功练习次数 N_T 的平均值, 分别为 1.48 次, 1.54 次和 1.55 次。Levene 检验的结果显示, 三组数据不满足方差齐性的条件 ($p=8.6\times 10^{-5}<0.01$)。Kruskal-Wallis 秩和检验的结果显示, 三组数据的均值具有显著性差异 ($p=2.5\times 10^{-4}<0.01$)。两两比较的 Nemenyi 检验结果显示, 声母组和韵母组具有显著性差异 ($p=5.3\times 10^{-3}<0.01$), 声母组和声调组具有显著性差异 ($p=4.9\times 10^{-4}<0.01$), 韵母组和声调组不具有显著性差异 ($p=0.85>0.1$)。

将所有 E_T 按时间顺序平均分为前期和后期两部分, 图 11 表示的是 N_T 的变化, 前期为 1.56 次, 后期为 1.49 次。置换检验结果显示, 前期和后期的 N_T 具有显著性差异 ($p=7.1\times 10^{-7}<0.01$)。

4.3 讨论

从发音练习量来看, 练习次数和时长都反映出学生利用 APP 完成了比较多的发音练习, 其中 90% 以上的发音练习是在课后完成的。这说明 APP 能够帮助学生利用课后时间完成自主的发音练习, 并且达到了一定的练习量。同时, 课堂内的发音练习量反映出 APP 起到了辅助教师进行课堂教学的作用。

从总体上看,声、韵、调的平均纠音成功率约为 83.2%,说明在 APP 的辅助下,学生能够改善大部分音的发音质量。声、韵、调的平均纠音成功练习次数约为 1.5 次,即仅通过 1 至 2 次的练习,学生的发音就会有明显的改善。进一步的统计发现,纠音成功练习次数为 1 至 2 次的占到了纠音成功练习次数的 89.0%,整体上练习次数较少。这说明在大多数的情况下,发音偏误能够通过自主练习的方式比较容易地得到改善。

对于重复练习的词条,每个练习的平均时长约为 2.5 秒。纠音失败的练习次数平均为 3.2 次,即经过 3 至 4 次的重复练习,若偏误的发音没有得到明显的改善,学生就会结束该词条的练习。

从组间差异来看,纠音成功率从高到低依次为:声母>韵母>声调,纠音成功练习次数从低到高依次为:声母<韵母<声调。这样的结果说明声母的纠音效果较好,声调的纠音难度较大。

在所有的纠音练习中,后期的纠音成功练习次数显著小于前期的纠音成功练习次数。这种行为变化说明,学生通过纠音训练后,纠音的效率有明显提高。纠音的结果是心理认知和生理产出共同作用的结果,是语音能力的一种体现形式。因此,纠音效率的提高体现了学生发音能力的提高。

目前初步的实验结果仅从总体上分析了发音练习行为和纠音结果,在重复练习的纠音过程中,学生利用多媒体信息辅助练习发音的行为数据还有待进一步地分析,以详细考察影响学生正音效果的因素。学生语音习得的长效性还需要通过专门的测试来检验。

5. 结论

本文介绍了智能发音教学 APP——尔雅中文 APP 的功能模块,在此基础上,研究了智能语音教学工具在汉语作为第二语言教学中对提高学生发音质量的作用。初步的教学实验研究结果表明:在智能语音教学工具的辅助下,学生利用课后时间自主地完成了发音练习的任务,并且达到了一定的练习量;基于具体声、韵、调正误的发音反馈信息,有效地引导学生纠正了发音偏误,改善了发音质量。目前,正音反馈信息主要是从发音混淆的角度提示学习者存在发音偏误,今后的工作将结合发音偏误趋势,提供更加具体的发音反馈信息,以期获得更好的正音效果。

致谢:本研究受北京语言大学语言资源高精尖创新中心经费(KYR17005)及校级重大基础研究专项(16ZDJ03)(中央高校基本科研业务专项资金)以及北京语言大学研究生创新基金项目(18YCX006)资助。作者感谢北京语言大学智能语音习得技术实验室解焱陆、李伟、郝丽霞、冯晓莉、及转转、牡丹、蔡睿、郑宝英等多位同仁的卓越贡献,也向大力支持教学实验的北京语言大学汉语学院曹文、张浩、刘敬华、唐翠菊老师表示感谢。

参考文献

- Cao, W., & Zhang, J. (2009). The construction of a CAPL Chinese interlanguage corpus and its annotation. *Applied Linguistics*, 4, 122-131. [曹文, &张劲松. (2009). 面向计算机辅助正音的汉语中介语语音语料库的创制与标注. *语言文字应用*, 4, 122-131.]
- Chen, J. (2018). *A comparative study of Chinese and Japanese phonetics of mandarin initial /r/ and /l/* (Unpublished master's thesis). Beijing Language and Culture University, Beijing, China. [陈静萍. (2018). 中日被试普通话声母/r/、/l/的语音对比研究(硕士学位论文). 北京语言大学, 北京, 中国.]
- Feng, X., Sun, Y., Zhang, J., & Xie, Y. (2014). A Study on the long-term retention effects of Japanese C2L learners to distinguish Mandarin Chinese Tone 2 and Tone 3 after perceptual training. *International Symposium on Chinese Spoken Language Processing* (pp.599-603). Singapore City, Singapore.
- Gao, Y., Xie, Y., Cao, W., & Zhang, J. (2015). A Study on Robust Detection of Pronunciation Erroneous Tendency Based on Deep Neural Network. *Interspeech 2015: Conference of the International Speech Communication Association*. Dresden, Germany.
- Godwin-Jones, R. (2011). Emerging technologies: mobile APPs for language learning. *Language Learning & Technology*, 15(2), 2-11.
- Kim, K. H. S., Relkin, N. R., Lee, K. M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, 388(6638), 171-174.
- Lenneberg, E., H. (1967). Biological foundations of language. *Hospital Practice*, 2(12), 59-67.
- Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27, 184-202.
- Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: anatomical changes in the human brain. *Cortex*, 58, 301-324.
- Li, Y., Cao, C., Xie, Y., & Zhang, J. (2017). A study on training Japanese students to discriminate nasal finals in standard chinse. *The 14th National Conference on Man-Machine Speech Communication*, Lianyungang, China. [李艳萍, 曹冲, 解焱陆, 张劲松. (2017). 日本学生汉语普通话鼻音韵母知觉训练初探. 第十四届全国人机语音通讯学术会议, 连云港, 中国.]
- Lin, J., Li, W., Gao, Y., Xie, Y., Chen, N. F., & Siniscalchi, M., ... Lee, C. H. (2018). Improving mandarin tone recognition based on DNN by combining acoustic and articulatory features using extended recognition networks. *Journal of Signal Processing Systems*, 90(7), 1077-1087.
- Liu, S., & Hung, P. (2016). Teaching pronunciation with computer assisted pronunciation instruction in a technological university. *Universal Journal of Educational Research*, 4(9), 1939-1943.
- Liu, Y. (2018). *A study of application of APP Chinese teaching software in extracurricular autonomous learning* (Unpublished master's thesis). Liaoning Normal University, Dalian, China. [刘洋. (2018). APP 汉语辅助教学软件在课外自主学习中的应用研究(硕士学位论文). 辽宁师范大学, 大连, 中国.]
- Mei, L. (2005). Variability in pronouncing Chinese blade-palatal initials by Japanese SLA

- learners. *Chinese Teaching in the World*, 1, 97-105. [梅丽. (2005). 日本学习者习得普通话卷舌声母的语音变异研究. *世界汉语教学*, 1, 97-105.]
- Mei, L. (2014). Cross-language speech perception in adults. *Journal of South China Normal University (Social Science Edition)*, 4, 164-169. [梅丽. (2014). 成人的跨语言言语知觉. *华南师范大学学报(社会科学版)*, 4, 164-169.]
- Neri, A., Cucchiari, C., & Strik, H. (2002). Feedback in computer assisted pronunciation training: Technology push or demand pull? *International Conference on Spoken Language Processing 2002* (pp. 1209-1212). Denver, Colorado, US.
- Qu, L., Xie, Y., & Zhang, J. (2017). A study of articulatory features based detection of pronunciation erroneous tendency. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 53(2), 239-246. [屈乐园, 解焱陆, & 张劲松. (2017). 基于发音特征的发音偏误趋势检测研究. *北京大学学报(自然科学版)*, 53(2), 239-246.]
- Krashen, S (1985). *The input hypothesis: Issues and implications*. London: Longman.
- Sun, Y., Zhang, J., Xie, Y., & Cao, W. (2013). Perceptual training of Japanese students to discriminate mandarin tones 2 and 3. *Journal of Tsinghua University*, 53(6), 921-924. [孙悦, 张劲松, 解焱陆, 曹文. (2013). 日本学生汉语阳平和上声的知觉训练效果的初步分析. *清华大学学报(自然科学版)*, 53(6), 921-924.]
- Wang, M. (2011). An analysis on the production of Chinese monophthongs by the Indonesian and Korean students. *TCSOL Studies*, 4, 16-25. [王茂林. (2011). 印尼、韩国留学生汉语单元音韵母发音分析. *华文教学与研究*, 4, 16-25.]
- Wang, Y., & Shangguan, X. (2004). How Japanese learners of Chinese process the aspirated and unaspirated consonants in standard Chinese. *Chinese Teaching in the World*, 3, 54-66. [王韞佳, & 上官雪娜. (2004). 日本学习者对汉语普通话不送气/送气辅音的加工. *世界汉语教学*, 3, 54-66.]
- Wei, X., Chen, J., Wang, W., Xie, Y., & Zhang, J. (2017). A study of automatic annotation of PETs with articulatory features. *Asia-Pacific Signal and Information Processing Association Summit and Conference* (pp.1608-1612). Kuala Lumpur, Malaysia.
- Wu, M., & Hu, M. (2004). An analysis of Vietnamese students' errors in Chinese tones. *Chinese Teaching in the World*, 2, 81-87. [吴门吉, 胡明光. (2004). 越南学习汉语声调偏误溯因. *世界汉语教学*, 2, 81-87.]
- Yang, J. (2013). Mobile assisted language learning: review of the recent applications of emerging mobile technologies. *English Language Teaching*, 6(7), 19-24.
- Zhang, J., Gao, Y., & Xie, Y. (2016). Mispronunciation tendency detection using deep neural networks. *Journal of Tsinghua University*, 56(11), 1220-1225. [张劲松, 高迎明, & 解焱陆. (2016). 基于 DNN 的发音偏误趋势检测. *清华大学学报(自然科学版)*, 56(11), 1220-1225.]
- Zhang, L. (2009). Perceptual training and the acquisition of Chinese aspirated/unaspirated consonants by Japanese students. *Language Teaching & Linguistic Studies*, 4, 85-90. [张林军. (2009). 知觉训练和日本留学生汉语辅音送气/不送气特征的习得. *语言教学与研究*, 4, 85-90.]
- Zhang, L. (2015). The contributions of segmental and prosodic information to the perception of foreign accent and speech intelligibility. *Chinese Teaching in the*

World, 2, 242-249. [张林军. (2015). 音段和韵律信息对汉语外语口音感知及言语可懂度的影响. *世界汉语教学*, 2, 242-249.