

基于 AI 智能语音技术线上教学的创新与实践

潘梦鸫, 吕小勇, 陈少伟, 邝锐铁, 王锋

(广东工贸职业技术学院汽车工程学院, 广东广州 510510)

摘要: AI 智能语音技术在教育教学中的应用重点是线上教学, 针对目前线上教学采用语音录播方式, 存在建设效率低、语音准确性低、录播时间长、录播受情绪和环境影响、线上课程建设质量落后等问题, 提出基于 AI 智能语音技术线上教学的创新方法和实践路径, 通过在线上教学应用智能语音技术, 提升教师个性化线上教学能力, 提高教师线上课程建设效率效果, 降低建设成本, 提高线上课程的个性化建设和课程建设质量。实践表明, AI 智能语音技术课程建设效率比传统录播课程建设效率高 10 倍以上, 该创新方法可全面提升高校教师线上教学能力, 促进 5G+AI 人工智能教学在高等院校的应用, 在提升高校教师教学的标准化教学水平、师资力量等方面发挥重要的作用, 可促进教育教学变革。

关键词: 人工智能; 智能语音; 语音合成; 语音转换; 线上教学; 辅助教学

中图分类号: G434

文献标识码: A

文章编号: 2096-5206(2022)12(b)-0170-04

Innovation and Practice in Online Teaching Based on AI Intelligent Voice Technology

PAN Mengyao, LYU Xiaoyong, CHEN Shaowei, HUAN Ruitie, WANG Feng

(School of Automotive Engineering, Guangdong Polytechnic of Industry and Commerce, Guangzhou Guangdong, 510510, China)

Abstract: The focus of implementation of AI intelligent voice technology in education and teaching lies on online teaching. At present, online teaching adopts the method of voice recording and broadcasting, with many problems existing, such as low construction efficiency, low voice accuracy, long recording and broadcasting duration, emotional and environmental influences on recording and broadcasting, and poor construction quality of online courses. This research proposes an innovative method and practical path for online teaching based on AI intelligent voice technology. Through the application of intelligent voice technology in online teaching, it is possible to enhance teachers' personalized online teaching abilities, improve the efficiency and results of teachers' online-course construction, reduce construction costs, and boost the personalized construction of online courses and the quality of course construction. Practice has shown that the efficiency of course construction by AI intelligent voice technology is more than 10 times higher than that of traditional course construction via recording and broadcasting. This innovative method can play an important role in comprehensively improving the online teaching ability of college teachers, promoting the development of 5G+AI teaching in colleges and universities, raising the standardized teaching levels of college teachers, and freeing teaching workforce, while delivering changes to education and teaching.

Key words: Artificial intelligence; Intelligent voice; Speech synthesis; Voice conversion; Online teaching; Auxiliary teaching

近年来,人工智能技术取得了飞速发展,智能语音技术为教育教学的发展提供了强有力的支撑。2017

基金项目: 广东省高等教育学会“十四五”规划 2021 年度高等教育研究课题“5G+AI 智能语音技术提升高校教师线上教学能力及路径研究”; 2021 年广东省高职教育教学改革研究与实践项目“基于 AI 智能语音技术线上教学的创新与实践”; 2021 年广东省普通高校创新团队项目“汽车行驶安全性能智能监控新技术创新团队”。

作者简介: 潘梦鸫(1970-), 男, 广东龙门人, 博士, 教授级高级工程师, 研究方向: 汽车新技术。

年,国务院办公厅印发《新一代人工智能发展规划》,指出人工智能的发展将会给人类社会和世界带来深刻改变,语音识别技术初步具备跨越发展的能力。2017 年,工业和信息化部印发《促进新一代人工智能产业发展三年行动计划(2018-2020 年)》,指出基于创新应用的语音识别、智能对话、语音合成、音视频融合将得到大力支持。

在新冠肺炎疫情防控期间,各大高校的线上教学得到了突飞猛进的发展,大部分线上课程建设采用语音录播方式,建设效率低、语音准确性低、录播时间长、录播受情绪和环境影响,导致线上课程建设质量

不高^[1]。人的语音包括音色、音高、音强以及音长四要素。音色是指听觉感觉到的声音的特色,由声音频谱决定;音高是指音的高度,即各种音调高低不同的声音,由机械波的频率和波长决定;音强是指声音的大小,由声波振幅大小决定;音长是指声音的长短,由发音体振动时间决定。语音四要素由不同语音特征参数共同决定,语音四要素又决定语音特性,导致人们听到不同(个性化)语音;说话人语音特征参数包括声学特征、韵律特征和语言特征。说话人语音特征参数包括声学特征、韵律特征和语言特征。声学特征是指说话人语音的音色,包括基频频率、共振峰位置、共振峰带宽、共振峰强度、频谱倾斜等;韵律特征是指说话的方式,包括音素的时间长短、音调和重音等(说话速率、音调和音量的变化);语言特征包括选词、方言和口音。AI 智能语音技术包括语音识别、语音合成、语音转换、语音编码等语音技术,是语音学与数字信号处理技术相结合的交叉学科,涉及语言学、信息学、脑神经科学、计算机学、心理学等多门学科。语音识别是电脑系统通过语音信号处理和模式识别自动识别和理解人类口述语言;语音合成是将输入的文本类型的信号序列经过适当的韵律处理后,通过特定的合成器,产生出高自然度、高音质、表现力丰富的语音输出,使计算机系统产生“人”一样自然流利的声音技术。个性化语音合成技术是对不同人的语音特征参数进行分析,提取说话人语音特征参数,用于语音合成系统进行个性化语音合成。语音克隆技术是指使用和训练一个新模型来完成对新发音人声音的克隆。语音克隆技术难点包括自然度、克隆语音与新发音人声音相似度;语音转换是指对一个说话人(源说话人)的语音中所包含的某些个性特征信息参数进行转换,使转换后的语音在语音语义信息不变的前提下具有另外一个说话人(目标说话人)的语音个性特征信息。语音合成质量判断标准:一是清晰度,是指语音内容清晰易懂程度;二是自然度,表示产生声音与人类语言相似程度。AI 人工智能教学即将机器人与人工智能、虚拟现实、多媒体等技术相结合,由机器人组织、实施与管理教学,优化教学结构、改进教学策略^[2-3]。

AI 智能语音技术是人工智能技术的基础之一,未来 AI 智能语音技术在研究上朝着语音识别、语音合成、语音交互技术方向发展,在应用上朝着智能家居、汽车、教育、医疗、安防和机器人等行业,其中,智能教育又是重中之重,AI 智能语音技术在教育教学中的应用重点是线上教学及语音交互辅助教学;AI 智能语音技术在线上教学的应用将大大提高教师线上课程建设效率、降低建设成本,提高线上课程的个性化建设和课程建设质量,全面提升高校教师线上

教学能力,同时,利用 5G 移动通信技术、云端大数据技术进行 5G+AI 人工智能教学,在提升高校教师教学的标准化教学水平、解放师资力量等方面发挥重要的作用。5G+AI 人工智能教学将为教育教学带来颠覆性变革,对教育教学的影响十分深远,具有广阔的市场应用前景^[4-6]。

AI 智能语音技术、5G 技术在线上教学的应用关系到高校广大教师教学能力的提高及教学方式的变革。国内外对智能语音技术的研究主要是围绕语音识别、语音合成、语音转换、语音交互 4 个方向开展,语音识别技术发展的时间长,已成为一项成熟、性能稳定的技术。对智能语音技术的研究现状分析,可以得到以下几个结论:(1)智能语音技术是提升高校教师线上教学能力的有效措施;(2)目前较少针对智能语音技术在个性化线上课程中的应用现状进行研究,教师线上课程的清晰度、自然度不够。

针对目前智能语音技术在教育教学行业应用研究现状所存在的问题,结合智能语音技术个性化、交互化等特点,本文提出基于 AI 智能语音技术的线上教学创新与实践,研究应用智能语音技术提升教师个性化线上教学能力及使用语音交互进行辅助教学,提高教师线上课程建设效率效果、降低建设成本,提高线上课程的个性化建设和课程建设质量,全面提升高校教师线上教学能力,促进 5G+AI 人工智能教学在提升高校教师的标准化教学水平、解放师资力量等方面发挥重要作用,为教育教学带来颠覆性变革。

1 AI 智能语音技术线上教学的创新方法

1.1 AI 智能语音技术线上教学创新方法

通过语音克隆技术克隆教师声音的教学主播进行个性化语音合成,可实现高自然度、高音质、表现力丰富的课件语音输出,通过语音转换,将课件源说话教师的个性特征语音课件转换为目标说话教师个性特征语音课件,通过人机语音交互、5G 移动通信技术实现辅助教学^[7],为实现 AI 人工智能教学提供方法和路径。

1.2 AI 智能语音技术线上教学创新路线

总体研究,制订出项目开发计划和需求分析,将系统功能划分成两个关键技术模块,确定各模块任务、分析模块间关系,实现目标分解,以便项目组更好地、协调地开展工作的。

(1) 建设教学主播语音合成系统。

通过语音克隆技术克隆教师的教学主播,将课件文本经过适当韵律处理,实现高自然度、高相似度、表现力丰富的课件语音输出,使教学主播像“教师本人”一样产生自然流利的语音课件。课件语音合成系统包括前端处理、韵律处理、合成单元^[8-11]。

(2) 建设课件语音转换系统。

将课件源说话教师的个性特征语音课件转换为目标说话教师个性特征语音课件,同时保持课件语义信息不变,保持教师的情感态度和自身声音特色,提高教师线上课程的真实度、自然度,营造身临其境的教学声音氛围。系统包括语音库、语音模型和参数、转换函数^[12-14]。

2 AI 智能语音技术线上教学的实践路径

2.1 实践研究

AI 智能语音技术线上教学系统由语音合成系统、语音转换系统两个部分组成。

2.1.1 准备

AI 教学主播工作室配置 AI 教学主播系统,可以进行教师样音克隆、语音合成、语音转换等工作。

2.1.2 克隆教师基准样音

教师在 AI 教学主播工作室使用声音克隆系统克隆教师基准样音。教师朗读声音训练文声音后,系统录制原始样音,原始样音不清晰可重新录制,直至清晰无杂音。

2.1.3 教学主播音库管理

教师原始样音经系统音频处理,形成教师的基准样音,称为教学主播播音人,添加到教学主播音库管理系统,供后期课件语音合成、转换调用。

2.1.4 线上教学课件音频合成

(1) 课件文本处理。以《智能网联汽车技术概论》课程为例,课程内容共分为 11 讲。

将《智能网联汽车技术概论》课件 PPT 内容、教师讲解内容,采用复制、粘贴、图形转文字等方式做前期处理,形成 Word 文字。文字一般从左至右,图片、表格内容按照学习者的学习顺序。用标点符号“,”“。”表示音频停顿时长,设置为不停顿、0.1 秒、0.2 秒、0.3 秒、0.4 秒、0.5 秒,《智能网联汽车技术概论》课程文字处理设置停顿时长为 0.3 秒。

(2) 课件教学音频合成。一是在 AI 教学主播系统调用系统样音库中的教学主播播音人;二是选择合适的播音人的播报语速、合成音调和合成音量参数;三是选择目标文字,复制到教学主播系统进行系统合成和储存,形成教学主播播音人的授课课件音频文件。

(3) 带背景音乐课件教学音频合成。为使课件教学音频文件更加生动、美化,把课件教学音频文件添加背景音乐,生成带有背景音乐的音频:一是添加选用的背景音乐;二是选择合适的人声音量、背景音量参数;三是进行带背景音乐的课件教学音频合成和储存。

2.1.5 线上教学课件视频合成

线上教学课件视频合成一般采用录屏方式,录屏

方式简单易行,速度快,但录屏时电脑系统运行声音对克隆音效会产生影响,如产生声音失真等现象。因此,录屏软件对电脑配置要求较高,以降低和消除系统运行对音效的影响。

在线上教学课件视频合成时候,先打开课件 PPT,同时播放课件音频文件,然后根据课件播放动作需要,选择不同的录屏制作软件合成和储存教学课件视频。根据制作需求选择录屏制作软件。

2.2 实践结果与分析

2.2.1 实践结果

采用 AI 教学主播课程建设《智能网联汽车技术概论》课程,共 46 段,合计 139 469 个字符。

2.2.2 实践分析

(1) 传统录播方法建设课程。

教师准备好所有的知识点材料,处理好前期文字,教师朗读录音速度为每秒 2.5 字符,则教师朗读时间为 626 分钟,假设教师因口误返工重读率为 50%,则教师朗读总时间为 $626 \times (1+50\%)=939$ 分钟,这是假设教师在录播过程中不间断、不休息、不耽搁情况下的计算时间。

(2) AI 智能语音技术方法建设课程。

AI 教学主播合成 2 000 字符约需要 1 分钟,AI 智能语音制作时间为 39.48 分钟,假设 AI 教学主播系统因网络故障、操作失误重录率为 25%,则 AI 智能语音制作时间为 $39.48 \times (1+25\%)=49.35$ 分钟。

工作效率计算: $939 \div 49.35=19.03$,则采用 AI 智能语音技术方法建设课程的工作效率为传统录播方法建设课程的工作效率的 19 倍,并且在传统录播过程中,教师语音准确性低、录播时间长、录播受情绪和环境影响,经常发生间断、休息、耽搁,采用 AI 智能语音技术方法建设课程的工作效率比传统录播方法建设课程的工作效率提高远大于 19 倍。

3 AI 智能语音技术线上教学的应用价值

3.1 具有较高的实践指导意义

研究成果为教师线上教学能力的提升和线上课程建设质量的提高提供了 AI 智能语音技术支持和实践指导。AI 智能语音技术在线上教学的应用将大大提高教师线上课程建设效率效果、降低建设成本,提高线上课程的个性化建设和课程建设质量,全面提升高校教师线上教学能力,同时,利用 5G 移动通信技术、云端大数据技术进行 5G+AI 人工智能教学,在提升高校教师教学的标准化教学水平、解放师资力量等方面发挥重要的作用。

3.2 具有较高的推广应用价值

研究成果在线上教学及人工智能辅助教学具有较高的推广应用价值,5G+AI 人工智能教学将为教育教

学带来颠覆性变革,对教育教学的影响十分深远,具有广阔的应用前景。

3.2.1 预期在校内进行推广

教学主播课件语音合成系统、课件语音转换系统研究成果预期在校内进行推广。

3.2.2 课题组成员单位进行推广

教学主播课件语音合成系统、课件语音转换系统研究成果预期在广东工贸职业技术学院、广东机电职业技术学院、广州科技职业技术大学等课题组成员单位进行推广。

3.2.3 预期向全省各大高校进行推广

教学主播课件语音合成系统、课件语音转换系统研究成果预期向全省各大高校进行推广,带动全省、全国各大高校 AI 智能语音线上教学及 AI 人工智能辅助教学的发展。

4 结语

(1) 建设教学主播课件语音合成系统。通过语音克隆技术克隆教师语音的教学主播高效输出教师语音课件,语音准确性高、输出不受情绪和环境影响,节省教师工作时间,大大提高教师课件制作的工作效率、降低建设成本。

(2) 建设课件语音转换系统。语音转换技术将源说话人个性特征语音的课件转换为目标说话人个性特征语音的课件,保持教师的情感态度和自身声音特色,提高教师线上课程的真实度、自然度,提高教学效果和教学质量。

参考文献

- [1] 修南. 高职院校线上“金课”建设思想理路与改进策略——基于数字原住民视角[J]. 中国职业技术教育, 2021(20):21-26.
- [2] 陈永平. 高职在线课程与课程体验融合教学设计: 运行状况、质量逻辑与提升途径[J]. 中国职业技术

教育, 2020(32):11-18.

- [3] 王雯, 韩锡斌. 面向远程技能训练的在线教学模式研究[J]. 中国职业技术教育, 2021(26):41-51.
- [4] 韦妙, 李朦. 人工智能对职业教育产教融合育人焦点及实践情境的形塑[J]. 中国职业技术教育, 2020(31):55-60,67.
- [5] 朱志萍. 智能释放: 人工智能 2.0 时代教育的冲击与改变——兼论人工智能赋能高等职业教育[J]. 中国职业技术教育, 2021(1):51-58.
- [6] 方益权, 黄云碧. 5G 时代高职教育发展范式的挑战与变革[J]. 中国职业技术教育, 2020(28):5-8.
- [7] 杨墨. AI 特定人语音及口型识别相结合技术[J]. 中国科技信息, 2020(21):88-89.
- [8] 陈小东, 宋文爱, 刘晓峰. 基于 LPCNet 的语音合成方法研究[J]. 计算机与数字工程, 2020,48(5): 1143-1147.
- [9] 王国梁, 陈梦楠, 陈蕾. 一种基于 Tacotron 2 的端到端中文语音合成方案[J]. 华东师范大学学报(自然科学版), 2019(7):111-119.
- [10] 热衣扎·哈那提, 努尔布力. 语音合成技术研究现状与发展趋势的计量分析[J]. 现代电子技术, 2019,42(21):116-119.
- [11] 张雄伟, 苗晓孔, 曾韵, 等. 语音转换技术研究现状及展望[J]. 数据采集与处理, 2019,34(5):753-770.
- [12] 王民, 苏利博, 王稚慧, 等. 采用 STRAGHT 模型和深度信念网络的语音转换方法[J]. 计算机工程与科学, 2018(9):1950-1954.
- [13] 李燕萍, 曹盼, 石杨, 等. 非平行文本下基于变分自编码器和辅助分类器生成对抗网络的语音转换[J]. 声学学报, 2020,59(3):322-329.
- [14] 曾歆, 张雄伟, 孙蒙, 等. 基于 GMM 模型和 LPC-MFCC 联合特征的声道谱转换研究[J]. 声学技术, 2020,39(4):451-455.

(本文编辑: 郭海峰)

(上接第 140 页)

- [10] 朱齐媛, 孟祥丽. 基于思想政治理念的“电子技术基础”课程教学改革探讨[J]. 科技与创新, 2021(13):67-68.
- [11] 顾佩华, 胡文龙, 陆小华, 等. 从 CDIO 在中国到中国的 CDIO: 发展路径、产生的影响及其原因研究[J]. 高等工程教育研究, 2017(1):24-43.
- [12] 韩新风, 高海涛. 基于 OBE-CDIO 模式的《数字电子技术》课程改革研究[J]. 现代计算机, 2020(20):74-79.

- [13] 单明广, 钟志, 于蕾. OBE 理念下的教学改革研究——以电子技术类课程为例[J]. 工业和信息化教育, 2021(3):51-54.
- [14] 曾凡菊, 谭永前. 基于 OBE 的数字电子技术教学改革[J]. 大学, 2021(39):122-124.
- [15] 王波, 王美玲, 刘伟, 等. 基于 OBE 理念的电子技术实践教学改革[J]. 实验室研究与探索, 2019,38(10):228-231.

(本文编辑: 郭海峰)