

人工智能技术在语音交互领域的探索与应用

陈胜森

肯特智能技术(深圳)股份有限公司 广东 深圳 518000

[摘要]在人工智能、5G技术等科学技术推陈出新的背景下,基于算法与信息汇集的智能设备改变了民众的生活方式,智能传播逐渐成为信息传递的重要形态,并伴随着技术的发展而不断扩张。于此同时,智能语音交互技术逐渐成为新时代人机交互的重要方式,并逐步成为智能设备与人类语言通信纽带的全新范式。当前,智能语音交互技术在民众生活中的渗透度逐步提升。伴随着产业与技术的蓬勃发展,智能语音交互技术的短板日渐突出。对此,本文将对人工智能技术在语音交互领域中的应用进行探析,用以强化对人工智能技术在语音交互领域应用的认知,提升人工智能与语音交互领域的衔接度,进而达到推动智能语音交互技术的优化发展。

[关键词]人工智能技术;语音交互;探索;应用

[DOI] 10.12252/j.issn.2096-6288.2021.09.262

人工智能技术在语音交互领域中的应用,即智能语音交互技术。目前,智能语音交互技术的概念界定与称呼尚未统一,使得智能语音交互技术常与“语音助手”、“语音识别”以及“人工智能”等概念混淆。整体惹眼,智能语音交互技术涉及语音识别、语言处理、人机对话以及语音合成等多项内容,能够有效提升人与机交流的顺畅度,满足更多人群的使用需求。对此,本文将从应用价值、应用技术以及具体化应用出发,对人工智能技术在语音交互领域的应用进行探析,借此强化对智能语音交互技术的认知,为智能语音交互技术的推广应用提供助力,为生活便捷性的提升创造条件。

1、人工智能技术在语音交互领域的应用价值

1.1外部助推力

人工智能技术应用于语音交互领域的外推力,主要表现为:一是传感技术的发展与推广,在传感技术不断发展的背景下,机器人对外感知能力逐步提高,具体表现在声音、温度以及速度等方面,为人工智能技术在语音交互领域中的提供了方向;二是通讯技术的快速发展,通信技术的发展为机器间的协同工作提供了重要支撑,让机器能够接收来自系统内部的指令,具体表现为导航软件以云端指令为依据播报信息;三是硬件技术的拓展与发展,当前,芯片技术的提高,推动了嵌入式设备运算能力的提升,相较而言,终端设备的功能性逐步提高,于此同时,GPU硬件性能的提升,亦为大规模学习计算任务的实现提供了可能,为人工智能技术在语音交互领域的应用提供了方向;四是大数据的发展,大数据技术与深度学习技术的提升,推动了文本、图像以及语音技术的发展,于此同时,在大数据的支撑下,推动了词义理解、语音识别效果的提升,

为智能语音交互技术的实现创造了条件。

1.2内部助推力

在传统人机交互中,基本流程为:前段设备接收→本地或云端语音识别(ASR)→自然文本转化→文本语义理解(NLU)→回答文本生成(NLU)→语音合成→客户端播放,具体流程如图一。

该人机语音交互模式的链路较为简单,演示级别的系统即可以满足运转要求。但是,在实际用户场景过程中,易遇到语音识别不准、语义理解不明、系统相应单一以及信息内容不充分等问题,具体表现为:一是语音识别不准,随着深度学习技术的不断提升,语音识别技术有所提升,且在用户配合的情况下,语音识别的可用性有所提升,但是,若出现口音、远近差异以及个性化词汇的情况下,极易出现识别不够理想的问题。而通过将人工智能技术的引入,能够提升语音识别的准确性;二是语义理解偏差,语义理解是语音交互的关键,而人类语言含有常识背景、场景特定语、语境以及口语化等多项元素,场景与语境等方面的切换增加了语义理解的难度,而人工智能技术的引入,有效提升了语义理解的分析能力,增加了语义理解的准确性;三是系统响应单一,以往设备反馈多以语言生成策略与对话策略为支撑生成反馈内容,反馈语言变化较少,且多以固定风格播报,风格变化的缺失,极易造成听觉疲劳。人工智能的引入,提升了反馈内容的丰富性;四是信息内容不充分,由于存储能力的限制,使得传统语音交互需要借助云端提供信息内容,而外部信息的授权机制会直接影响交互的顺畅性。人工智能技术的引入,提升了信息储存能力^[1]。

1.3人工智能语音交互技术的具体化应用

当前,智能语音交互仍属于新型技术,在市场拓展方面仍处于初级阶段,但是,伴随着技术的不断拓展,智能语音交互技术的应用范围逐步拓宽(如图二),具体可以划分为企业级、工业级以及消费级等方面,其中消费级主要包含智能家居、智能教育以及智能商务等,而企业级主要包含智能客服、智能媒体以及智能公检法等,而工业级主要包含安防系统、智慧城市以及AI新基建等方面。

2、人工智能技术在语音交互领域中的应用技术

以人工智能技术的特点为支撑,语音交互是前后端结合、



图一 传统人机交互



图二 智能语音交互技术的应用范围

软硬件一体的过程，在此过程中，主要涉及声源定位、人声检测、音频采集、语音唤醒、语音听写以及语义理解等多个过程（如图三），其间设计的人工智能技术主要涉及以下几方面：



图三 人工智能下的语音交互流程

2.1 声学前端技术

在人机交互远场景下，易受到人声、噪音以及回响等声音的干扰，该问题多采用多麦克风阵列的方式进行解决，旨在通过多渠道采集音频与参考信号的方式，提升音频采集的准确性，于这一模块而言，需借助硬件结构调整来处理，减少因硬件结构而产生的噪音源，于此同时，再配合定向拾音与前段声学算法，处理回声、噪音所带来的影响。

在人机交互过程中，语音唤醒是对话触发的主要方式，语

音唤醒技术能够让休眠状态的设备进入待指令状态，属于语音交互开启的第一步。针对此，在接入人工智能技术时，应当注意误唤醒率与唤醒率的控制，尽可能保障在理想状态下的百分百回应^[2]。

此外，语音信号质量的高低还与语音方位、角度识别技术息息相关，定向拾音则是保障降噪处理准确性与人声增强的重要路径。定向拾音技术多借助多麦克风阵列、声源定位技术以及波束形成技术，对声源与麦克风阵列的距离与角度进行计算，并对目标声源进行跟踪，于此同时，在期望声源方以上，形成有效波束，仅接收波束内信号，达到提取声源与降噪的效果。但是，该方式下采集的音频仍带有背景噪声，语音降噪技术有效解决了该问题，但是，于不稳定脉冲噪音和欺负噪音而言，降噪效果更不够理想。

随着智能设备市场覆盖率的提高，远程语音交互在设备中的应用度逐步提升，其本身支持声音的输出，具体包含智能车载、智能影像等。但是，这类智能设备仍会遇到设备播放音与人声指令重叠的情况，直接影响识别效果，但是，借助回声消除技术，能够对扬声器声音进行消除，进而达到提升语音拾取质量的效果^[3]。

2.2 语音识别技术

远场拾音拥有了波束形成技术与声源定位技术，有效提高了语音拾取的质量，但是，仍有较多的无效语音与噪音被采集，增加了语音识别的难度，在此背景下，需要借助端点检测

技术对有效人声进行检测,并将非人声过滤而出,不仅能够降低系统乱响发生的可能性,还能有效降低语音识别服务与通信网络的负载。

一般而言,语音文本的获取多经由端点检测再到语音识别,且随着技术的发展,语音识别在英语口语、普通话乃至方言上皆取得了良好成果,特别是在前端技术的影响下,语音识别的可用性有所提高,为语音交互提供了基础保障。

但是,随着市场的发展变化,语音交互给语音识别提出了新要求,具体表现在特殊场景交互等方面,而部分智能设备在联网上有所局限,需要将语音识别与语义理解本地化,这给语音交互技术带来新挑战,特别是在强视觉语音交互终端中,需要界面所有元素的语音交互,此时,需要设计特定识别系统,用以保障界面元素的实时动态获取与生效,以提升各类信息从语言层面上的精准识别,即“所见即可说”^[4]。

2.3 对话理解技术

于语音识别而言,持续性音频流识别是基础,对此,在持续性识别语音中,极易吸收较多的无效语音,此时,需要借助拒识别技术对无关语音和噪音进行过滤,并在此基础上,借助上下文对话引擎对识别语音结果进行词义理解,在此过程中,词义理解不单需要对用户的说话意图进行识别,还需要对语音内容进行精确获取,以保障完整交互链路的形成。

一般而言,在用户意图理解过程中,需要对交互场景下的口语表达、实体歧义等进行识别,且意图理解的首要目标,涉及文本关键信息槽、交互意图识别等多项内容,其中信息槽位识别属于序列标注过程,较为典型的技术包含卷积神经网络、双向循环神经网络等多种模型,且皆已渗入深度学习技术,使其具备一定的成长空间。但是,由于特定对话训练数据不足,使得模型的语义理解能力偏弱,对此,应当借助实体、句式等资源进行填充,或通过提高语义建模能力的方式,促进泛化性能的提升^[5]。

除却用户意图理解,在对话理解引擎之中,还需涉及灵活对话管理模块。特别是在多轮对话之中,对话管理的价值不言而喻。一般而言,智能化的对话管理模块,需要融合存储的上下文信息、语义理解结果内容等数据信息,以保障系统响应的合理性。由此,稳定且持续的高质量内容信息源与数据库尤为重要,一般为对话系统的必备模块。此外,在对话系统还应当具备语义纠错能力,以降低意图理解失准与内容噪音过大而引发系统相应偏差问题出现的可能性。当前,在应用级对话系统之中,该模块多以人工专家设计的文法与规则来实现。

此外,在实际对话系统之中,功能模块亦十分重要,具体表现为闲聊对话、知识问答以及情感分析等功能,兼具多种回复能力的系统,能够有效提升人机交互的体验感。于此同时,图像、指纹以及手势等数据信息皆能够有效丰富交互对话能力,亦逐步成为对话系统的未来发展趋势^[6]。

2.4 语音合成技术

语音合成技术相当于机器的嘴巴,一般而言,语音交互多

由两个语音合成系统组合而成,即拼接语音合成与波形建模语音合成,借助拼接合成系统能够合成受限制的固定文本格式模板,且在限制领域内,语音表现具有较高的自然度与表现力,借助大量模块设计即能够促进交互体验感的提升,借助波形建模合成系统则能够合成动态化部分。此外,在交互设计过程中,通过在拼接合成系统中适度增加副语言偏度,即“哦”、“嗯哼”等词汇,能够有效提升对话的真实感。当前,波形建模语音合成的典型工作为wavenet,其具有合成音质高的优势,可以直接与拼接语音或原音进行组合,且听感上的差异偏小,但是,其存在过度依赖训练数据量的情况,且计算量易偏大,增加了运转负担^[7]。

2.5 端点检测技术

当前,端点检测多依赖于音频相对能量变化,若用户说话较为犹豫,易出现音频切断的情况,进而导致语音交互失败,从该角度出发,应当充分对人类语言的完整因素与持续性特定进行把握,若用户说话未完成,则需要增加等待时长,若说话完成,则缩短等待时长。在此过程中,语言完整度判断尤为重要,而综合判断技术即“语义端点检测”。在此过程中,应当借助AI技术提取个性化数据,以促进语音交互质量的提高。

结束语

综上所述,在市场需求的推动下,智能语音交互技术的应用范围逐步扩宽,于此同时,随着应用率的提升,语音识别不准、语义理解不明、系统相应单一以及信息内容不充分等问题逐渐显露而出。在此背景下,应当重视人工智能与语音交互的深度融合,通过优化声学前端技术、对话理解技术、语音合成技术、语音识别技术的方式,促进人工智能与语音交互融合度的提升,促进语音交互质量的提高。

参考文献

- [1] 刘志琳. 智能语音交互技术的传播特征研究[D]. 河北大学, 2021.
- [2] 刘宏, 郝乾啸, 萧嘉荣. 人工智能语音交互技术在电力调控领域的应用[J]. 电子技术与软件工程, 2021(04): 217-219.
- [3] 刘灵景. 基于用户使用与满足情况的智能音箱发展问题及对策研究[D]. 西北大学, 2019.
- [4] 李敏. 人工智能时代新闻伦理面临的困境及应对策略研究[D]. 辽宁大学, 2019.
- [5] 邝展鹏. 语音交互设计与研究[D]. 华南理工大学, 2019.
- [6] 陈志刚, 刘权. 人工智能技术在语音交互领域的探索与应用[J]. 信息技术与标准化, 2019(21): 16-20.
- [7] 高瑾. 语音互动广告的交互特性对用户广告接受度的影响研究[D]. 华南理工大学, 2018.

作者简介:

陈胜森(1974年7月2日)男,汉族,江西省赣州市人,本科,研究方向:智慧人居环境及服务,智能化控制系统。