

具身智文本内容生成的咖啡机器人系统设计^①

陈华舒^{②*} 卢振利^{③**} 王伟栋^{***}

(* 盐城工学院机械工程学院 盐城 224051)

(** 常熟理工学院电气与自动化工程学院 常熟 215500)

(*** 苏州东控自动化有限公司 常熟 215500)

摘要 本文基于东控协作机器人臂、咖啡机,融合智文本内容生成-生成式预训练转换器(AIGC-GPT)的大语言生成式模型设计了一款智能咖啡机器人系统,该系统能够实现基于 ChatGPT 的语音对话,根据语音对话自动化制备咖啡及咖啡拉花。通过 Azure 认知服务(ACS)技术整合协作机器人和 GPT,使咖啡机不再仅仅是一台传统的咖啡制作设备,而是一个集智能、互动、个性化为一体的咖啡制作平台。这种设计包括语音控制的自动咖啡制作与点单、个性化推荐、咖啡品质控制等。这种创新性的融合有望为咖啡业和其他领域带来新的机会,提高用户满意度和创造更具吸引力的咖啡馆体验。

关键词 协作机器人; ChatGPT; Azure 认知服务(ACS); 人机交互

传统咖啡需要人工操作特殊的设备制作咖啡,包括点单、研磨、冲泡、清理杯具等。这增加了人工成本,并会因为工作时间的延长,导致员工疲劳和产品品质不一致,影响咖啡的品质口感。所以大部分咖啡店为了缩短制备的时间,会利用咖啡机制作咖啡,这样可以保证产品质量的一致性。

虽然市面上大部分咖啡店都是利用咖啡机制备咖啡,但当顾客需要特殊的拉花咖啡时,就需要有经验的咖啡师制备,培养一名合格的拉花咖啡师成本较高,且拉花过程耗时较长。而协作机器人具有操作简单、方便灵活、安全可靠以及与其他先进技术易融合的特点^[1],因此本文创造性地提出了使用协作机器人与生成式预训练转换器(generative pre-trained transformer, GPT)融合的方式^[2],设计出一款可语音交互的智能咖啡机系统,使得顾客能够在点单的过程中和咖啡机器人展开日常交流与对话,增加点单过程中的趣味性。

2017 年 CAFE X 公司推出智能咖啡机器人^[3],

将自动冲煮技术、机器人技术、点单系统以及多家精品咖啡烘焙公司的产品结合在一起,大幅减少人工工作,但此咖啡机器人实现的功能较简单,机械臂主要负责端和拿的动作。2021 年,北京如影智能科技有限公司通过自主研发的 inBotOS 系统^[4],将优傲 ur3 型号协作机器人臂与高端 Eversys 咖啡机融合,上市了一款智能咖啡机器人,该机器人可以进行点单、收银、拉花、制作咖啡、冲洗等功能,但造价十分昂贵,并不具备语音互动功能。本文将智文本内容生成(artificial intelligence generated content, AIGC)嵌入智能咖啡机系统^[5],使其拥有强大的自然语言对话功能,可以与顾客无障碍的日常对话,再通过专属大语言模型的训练,使得该系统能够实现咖啡点单的对应操作。在点单咖啡的同时,也能使顾客体验到对话的乐趣。本系统采用了分布式架构^[6],由协作机器人、GPT4.0、个人电脑(personal computer, PC)端、咖啡机、工作台、打印机和扫码器组成。

① 中塞第 5 届科技例会交流项目(国科外 2021-4-19)资助。

② 男,2000 年生,硕士生;研究方向:机器人智能控制;E-mail: chs200005@hotmail.com。

③ 通信作者,E-mail: zhenlilu@cslg.edu.cn。

(收稿日期:2023-10-13)

1 咖啡机器人系统架构设计

硬件设备由东控公司的协作机器人、咖啡机及其配套设备、PC端、打印机、扫码器组成。PC端作为服务器,2台协作机器人和咖啡机分别作为下位机1、下位机2和下位机3,本文中接咖啡的协作机器人称为左机器人,接奶的协作机器人称为右机器人,协作机器人与上位机使用场以太网总线(EtherCAT)和TCP/IP以太网数据网络进行连接,咖啡机与上位机通过RS232接口转化,使用USB的COM口建立串口通讯。硬件连接逻辑图如图1所示。

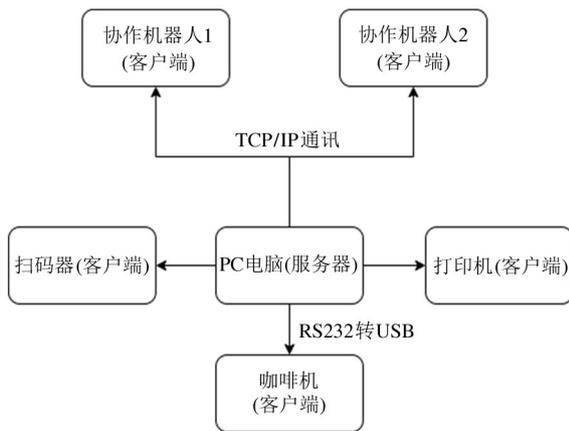


图1 硬件连接逻辑图

1.1 硬件设计方案

两台协作机器人通过东控公司的 rPAC-540 一体化控制器驱动控制,内置现场以太网总线(EtherCAT)和TCP/IP以太网数据网络,可以精准完成拉花动作。

控制器分别由RS232、RS485、USB、VGA等接口部分组成,支持PTP、LIN、CIRC示教编程,且支持开放式OEM客制化C/C++编程。

控制柜的主电源回路接线分别由主控开关、旋钮开关、风机、机器人航插、制动电阻板、IO模块、示教器、控制器组成,其逻辑关联图如图2所示。

其中,风机采用德力西轴流风机DZ47 C32 2P,规格为80 mm × 80 mm × 25 mm的直流24 V滚珠,转速为4 100 r · min⁻¹,在电气柜中主要用于24 V电源和48 V电源的散热。制动电阻版为800 W、15 Ω的安永电阻电子版,IO模块型号为SMART-IO-R1.1,用于连机器和设备,示教器型号为HT0803BV-1,机

器人控制器为东控公司自主研发的rPAC-560-OEM控制器。

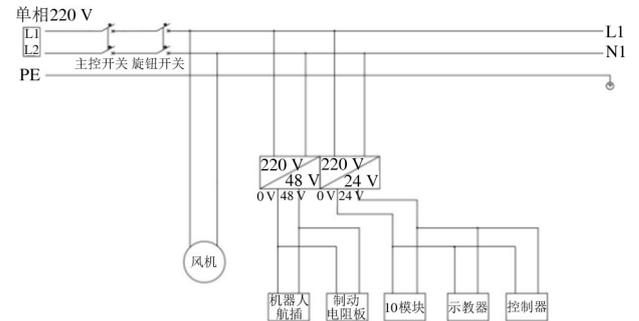


图2 控制柜主电源回路接线

协作机器人负载3 kg,臂展622.00 mm,重复定位精度误差为0.04 mm,咖啡机型号为瑞士Eversys Cameo品牌,可自动完成磨豆、冲泡、倒奶等工序,也可实现无人清洗机器人内部管道,扫码器扫描二维码,其组装后效果为图3所示。



图3 智能咖啡机硬件组装

1.2 软件设计方案

上位机的选用系统软件选用Win 11系统,咖啡机通过RS232转换器转换USB串口通讯,使用网口协议与2台协作机器人通讯,发送指令控制协作机器人和咖啡机。左协作机器人的IP设置为192.168.58.2,右协作机器人的IP设置为192.168.58.3,端口号均为2100,使用Python语言搭建整体的配置系统。上位机通过VS2020软件发送指令代码控制咖啡机与协作机器人1和2,相应指令与动作如表1所示。

PC端上位机将扫码器收到的数据转换成“2、3、4、5”发送给左、右协作机器人,机器人在经过程序判断后,在工作的过程中根据收到的状态发送下一步指令控制机器人继续进行工作,将串口接收的16位进制的复杂字符串简约成控制指令。

表 1 咖啡机动作对应表

指令字符	咖啡机动作
“1”	出咖啡
“2”	出奶
“3”	出美式咖啡

扫码器与上位机使用 RS232 通讯转 USB 连接,扫码器设置波特率为 115 200,将二维码的识读模块配置成虚拟串口输出模式,通过扫描拉花咖啡的二维码,得到相应编码,从而实现对应拉花咖啡的制作。

协作机器人通过网口使用 Python 的 socket 库进行网络通信,它与咖啡机通过网络套接字建立连接,发送和接收命令和数据。网络通信用于控制咖啡机的启动、制作咖啡等操作,通过 Socket 对象的 connect_ex()、send() 和 recv() 方法来实现连接远程主机、数据发送和数据接收,使得程序能够与远程主机进行数据交换,系统架构如图 4 所示。

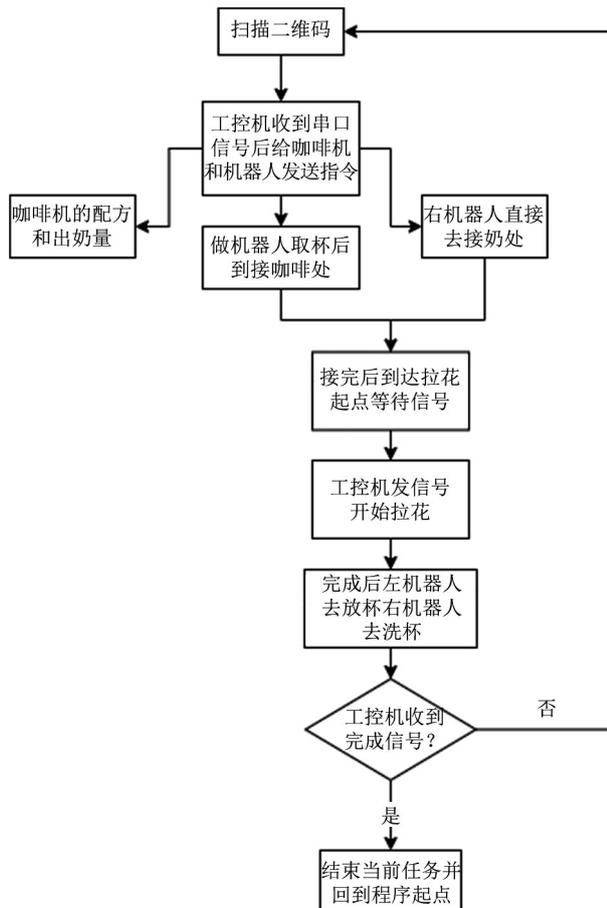


图 4 整体系统架构图

2 协作机器人拉花动作标准设计

协作机器人要实现末端轨迹的准确运行,需要精准的运动学模型,基于变换矩阵的建模方法又叫 D-H(Denavit-Hartenberg)法^[7],该方法可以表示各关节坐标系之间的相互变换,适用于各种机构的机器人运动学建模。根据运动学建模,导入协作机器人的示教编辑系统,通过程序编辑完成轨迹规划。

2.1 基于变换矩阵的运动学建模

在不考虑力的情况下,通过各关节所建立坐标系的变换关系来描述关节空间和笛卡尔空间之间的位姿关系。机器人运动学又分为正运动学和逆运动学^[8]。其中正运动学可以通过将各个坐标系的变换矩阵相乘,得到描述机器人末端位姿的变换矩阵;而逆运动学则需要对这个变换矩阵进行逆运算,以求解出各个关节的转动角度^[9]。

机器人每个连杆由 4 个参数来定义^[10],分别为 d 、 θ 、 a 和 α ,其中 d 表示该坐标系的 z 轴到下一个坐标系的 z 轴的距离, θ 表示该坐标系的 x 轴绕 z 轴旋转的角度, a 表示该坐标系的 x 轴到下一个坐标系的 x 轴的距离, α 表示该坐标系的 z 轴绕 x 轴旋转的角度。根据东控公司的产品说明书,可获得 D-H 参数模型^[11]。

机器人相邻坐标系下的位姿变换矩阵 ${}^{i-1}\mathbf{T}_i$ 可由式(1)所示的数学推导公式变换而来^[12]。

$${}^{i-1}\mathbf{T}_i = \text{Rot}(x, \alpha_{i-1}) \text{Tran}(a_{i-1}, 0, 0) \text{Rot}(z, \theta_i) \text{Tran}(0, 0, d_i) \quad (1)$$

由式(1)可推导出相邻 2 个坐标系之间的齐次变换通用矩阵为

$${}^{i-1}\mathbf{T}_i = \begin{bmatrix} \cos\theta_i & -\sin\theta_i & 0 & 0 \\ \sin\theta_i & \cos\theta_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & a_{i-1} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha_i & -\sin\alpha_i & 0 \\ 0 & \sin\alpha_i & \cos\alpha_i & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i \cos\alpha_{i-1} & \cos\theta_i \cos\alpha_{i-1} \\ \sin\theta_i \sin\alpha_{i-1} & \cos\theta_i \sin\alpha_{i-1} \\ 0 & 0 \\ 0 & a_{i-1} \\ -\sin\alpha_{i-1} & -d_i \sin\alpha_{i-1} \\ \cos\alpha_{i-1} & d_i \cos\alpha_{i-1} \\ 0 & 1 \end{bmatrix} \quad (2)$$

2.2 拉花轨迹规划设定

更改接入电脑的 IP 地址,通过 TCP/IP 协议与协作机器人建立通讯,登入东控协作机器人的控制台,如图 5 所示,依赖运动学推导的通用矩阵数学模型,设置好初始的标准工具坐标系。

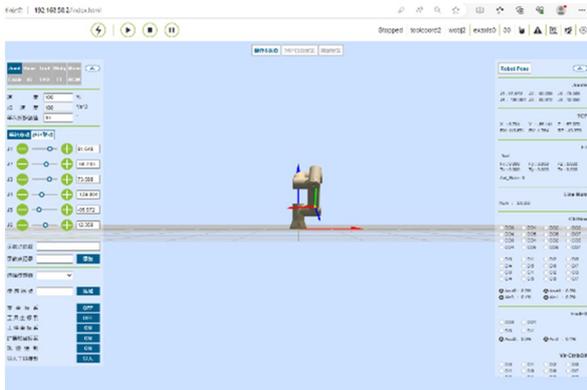


图 5 协作机器人示教界面

进入拖动示教中的程序示教,设置完成初始姿态后计算关节位置,将右机器人移至准备拉花的位置,如图 6 所示,通过东控协作机器人的指令调用,完成最后的拉花咖啡,如图 7 所示。

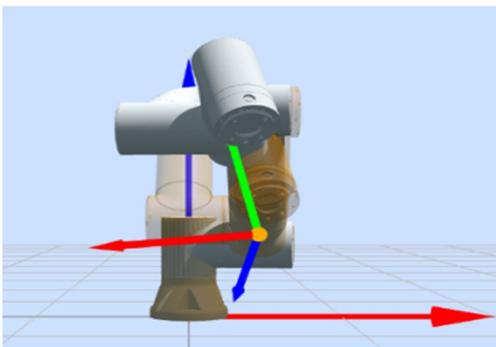


图 6 协作机器人拉花初始姿态设定



图 7 最终拉花效果

3 基于 AIGC 的语音交互模块

语音交互系统模块分别由 4 个主要部分组成,分别为人脸识别、AIGC-GPT 的文本交互处理、文本音频交换处理和命令指令打印与扫描^[13]; OpenCV 捕获来自摄像头的图像^[14],触发人脸识别,通过 Azure 认知服务 (Azure cognitive services, ACS) 的语音识别器,输入对话,并利用 speech_recognizer 进行处理转化为文本,调用 OpenAI 的 API,发送到 AIGC-GPT 模型,生成自然语言^[15]。当提到“咖啡”关键词后,会进入点单环节,根据顾客要求,输出对应拉花咖啡的文本,通过调用 print_docx 函数,打印出相应咖啡二维码。

3.1 人脸识别分析

本文通过 OpenCV 捕获来自计算机的摄像头中的图像帧,对图像进行预处理,将捕获到的图像传入 deepface 库,该 deepface 库基于 vscode 平台搭建卷积神经网络模型^[16],识别和分析面部特征,包括性别、年龄,通过分析大量图像人物来确保模型能够学习到不同性别的面部特征。

整体结构共由 4 层组成,分别为卷积层、全连接层、激活函数以及分类层。

实验对象来自网络数据库,卷积层提取图像特征,是通过卷积操作检测局部的视觉模式,参数设置为 32 个卷积核卷积核大小为 (11, 11); 池化层可以减小特征图的尺寸,保留主要的特征并降低计算复杂度,池化窗口大小为 (3, 3),步幅为 2,使用“same”填充;全连接层将卷积层提取的特征映射到输出类,在 deepface 设置中一共有 4 096 个神经元学习图像;最后,在分类层中通过全连接层输出,对

输入的人脸图像进行分类,使用 softmax 激活函数,将模型的输出转化为分类概率,经过训练,达到识别性别的目的。

deepface 通过以上处理,可判断性别,提示音响起,进入与 GPT 的语音交互。流程图如图 8 所示,效果图如图 9 所示。

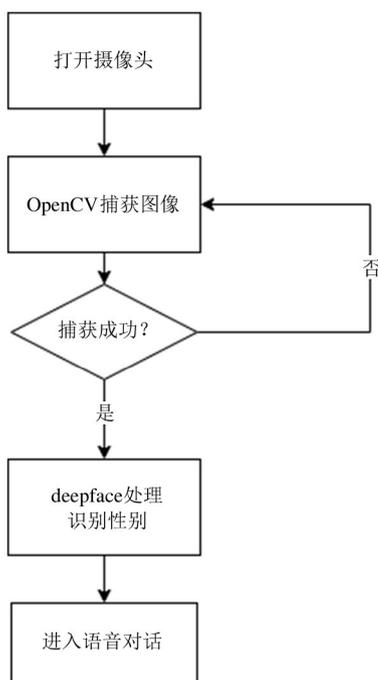


图 8 人脸识别流程图



图 9 人脸识别效果图

3.2 基于 AIGC 的文本交互处理与专属大模型训练

结合 AIGC 和大语言模型 (large language model, LLM)^[17],协作机械人可以实现人工智能思维和行动操作的统一。本文提出了 AIGC 与机器人融合的技术设计路线图,如图 10 所示。路线图的关键部分是训练一个专属的大语言模型 LLM,即针对机器人领域特定需求。

使用 azureOpenAI.chat 函数,与 AIGC-GPT 语言模型进行交互,以生成响应用户输入的内容。这个函数会将用户输入发送给语言模型,并获取响应,建立专属大模型训练,训练与咖啡指令相关的语言



图 10 AIGC 模块思路导图

模型。如图 11 所示,当用户只是想与上位机进行普通的文本交互时,该系统直接调用 AIGC-GPT 与用户交流,当提到“咖啡点单”等关键词时,系统会通过 template.chat 函数,包括模板文本,用于指导或补充与用户的对话。

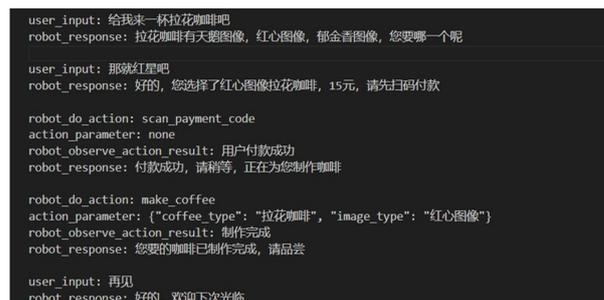


图 11 AIGC 训练效果图

3.3 文本音频交换处理

当进入对话环节,用户通过 PC 端自带麦克风输入语音,通过 ACS 中的 speech 模块实现语音识别。如图 11 所示,通过提供 API 密钥和端点信息,程序可以将用户的语音输入发送到 Azure 云服务,然后将其转换为文本,通过 if 语句,判断关键变量 cap = VideoCapture 是否为 0,如果是 0,则判定用户可自由与 GPT 实现自然语言对话之后,通过 ASC 中的 synthesizer 函数将文本转化为语音,当提到“我要一杯咖啡”或者“请给我来一杯咖啡”等与点单相关的关键词后,会进入点单环节,如图 12 所示。



图 12 对话点单效果

3.4 命令数据打印与扫描处理

当顾客确认需要哪一种咖啡时,通过 print _

docx 函数生成并打印 Word 文档,Word 文档中设置好相应咖啡的二维码,提前设定好关键变量,将变量和相关 word 文档一一对应,该模块整体流程如图 13所示。

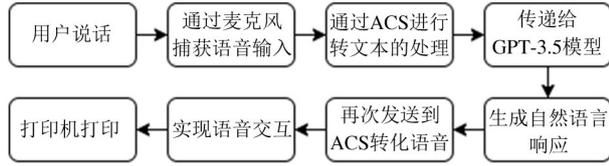


图 13 语音交互流程

在调用打印机时,使用了 win32com 库中的 COM 对象来与 Microsoft Word 应用程序进行交互,打印指定的 Word 文档,通过调用 print_docx 函数来打印指定的 Word 文档,具体的文档文件由 coffee 参数决定。打印出的二维码和扫描器如图 14 所示。



图 14 二维码打印与扫码器

在获得对应拉花咖啡的二维码后,将二维码贴在纸杯上,扫描扫码器,并放在相应位置,协作机器人与咖啡机自动运行相应程序。

4 上位机人机交互界面设计

人机交互界面选用 Python 编程软件中的 thinter 库进行程序设计,界面有 4 种咖啡按钮,用于不同标签的交互和处理,可选语音点单,也可选择手动点单,如图 15 所示。

使用 create_sub_window 函数用于创建系统设置的子窗口,其中包括 4 个勾选框,用于选择不同类型的咖啡,当咖啡豆或者牛奶不足时,可在设置中设

置暂时缺货的咖啡,如图 16 所示,这时候就不会打印出相应咖啡的二维码。



图 15 人机交互界面



图 16 咖啡设置界面

按钮通过 threading 函数导入 Python 的多线程库,使用 beepThread.start() 启动 beepThread 线程。这将使线程开始执行 beep 函数中的操作,并且可以与主线程并发运行,当顾客想使用手动点单时,可通过人机交互界面进行手动点单。

5 系统实验测试与结果分析

为了测试该系统的鲁棒性和实时性,本文利用该智能咖啡机进行 50 杯咖啡的制作,实验相关要求如下。

5.1 实验场地布置与方案设计

测试者一共分为 5 组,每组 1 人,分别点单 10 次。测试过程中设备视角内尽量只能有测试者 1 人,出现多人时会影响系统识别准确度。此外,如果现场比较嘈杂,在测试过程中系统识别的图像和声

音会较杂乱,从而影响图像识别的声音准确率和声音录入的质量。测试者在设置的模拟复杂情况下通过语音交流点单,分别从系统识别人脸次数、完整进行对话情况次数 2 个方面记录 50 次咖啡点单的情况。

5.2 咖啡机制作流程设计

该系统样机在正式测试时,需提前打开咖啡机进行锅炉预热,测试者能够正常被识别人脸的合适距离,该过程 5 min 之内即可完成。在找到合适识别人脸范围后,运行该系统程序,系统运行后检查人脸识别程序是否有报错情况,在无报错情况下开始进行正常测试。

5.3 测试结果分析

通过测试实验,在不同情况下对系统识别到人脸次数、能够进行完整对话情况次数、智能咖啡机制作运行时间与传统制作所需时间进行数据对比分析,绘制成折线统计图,如图 17 和 18 所示。

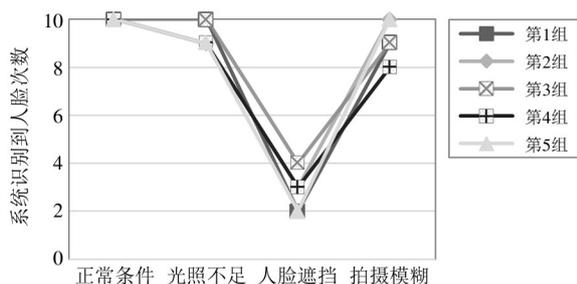


图 17 不同情况下系统识别到人脸次数

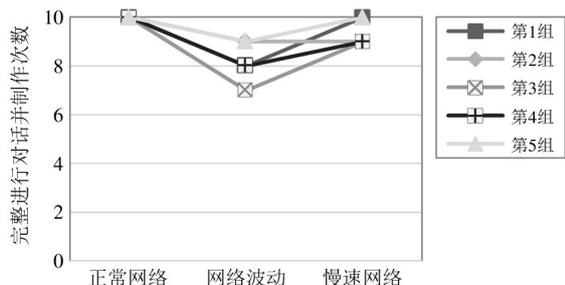


图 18 不同情况下完整进行对话情况次数

50 杯咖啡的制作过程中,正常条件下咖啡全部可以制作完成,并且被测者能够与系统展开正常对话并输出语言文本,相关对话如图 19 所示;在人脸遮挡和网络波动的不良条件下,会导致人脸识别失败和语音卡顿,最终导致点单失败,但咖啡机器人系

统在网络恢复正常时可以继续进行正常对话聊天。

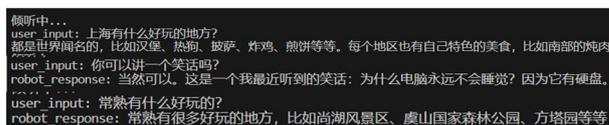


图 19 语音对话效果展示

6 结论

该技术将咖啡机、AIGC-GPT 与协作机器人结合,在大模型训练的基础上,仅通过顾客的语音对话,实现无人情况下的咖啡点单,减少咖啡售卖的人工成本,提高点单过程中的趣味性。但本系统需要连接互联网,在对话过程中可能因为网络波动导致卡顿等现象,后续将研发本地化部署问题。本研究针对无语音控制协作机器人的空白,设计了基于 AIGC-GPT 的一款可人机语音交互的智能咖啡机,为其他相关 AIGC 应用提供了技术储备。

参考文献

- [1] 刘华秋,黄磊,陈逸维. 协作机器人国内外发展现状与技术研究[J]. 现代制造技术与装备, 2023, 59(3): 93-96.
- [2] 张熙,杨小汕,徐常胜. ChatGPT 及生成式人工智能现状及未来发展方向[J]. 中国科学基金, 2023, 37(5): 743-750.
- [3] HU M, WANG H, PAN X. An extended stiffness model for 7 DoFs collaborative robots using the virtual joint method[C]//2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). Dali, China: IEEE, 2019:1653-1658.
- [4] 史玉红,蔡丽娟,蔡佳丽,等. 协作机器人在智能制造中的应用[J]. 集成电路应用, 2022, 39(2): 196-197.
- [5] 王全,解迎刚,杨海林. 协作机器人智能视觉系统设计与实现[J]. 物联网技术, 2022, 12(2): 28-29.
- [6] 朱永飞. 双臂协作机器人避障路径规划与控制研究[D]. 马鞍山: 安徽工业大学, 2021.
- [7] 李海旺,隋春平,王宇坤,等. 协作机器人运动学应用标定方法[J]. 中国科学院大学学报, 2023, 40(4): 555-565.
- [8] 田志伟. 协作机器人无传感器碰撞检测方法研究[D]. 天津: 天津大学, 2018.

- [9] HOGAN N. Impedance control-an approach to manipulation: part I-theory [J]. Journal of Dynamic Systems Measurement and Control-Transactions of the ASME, 1985,107(1):1-7.
- [10] MORGAN A S, HANG K Y, BIRCHER W G, et al. Benchmarking cluttered robot pick-and-place manipulation with the box and blocks test[J]. IEEE Robotics and Automation Letters, 2020,5(2):454-461.
- [11] STENMARK M, HAAGE M, TOPP E A, et al. Supporting semantic capture during kinesthetic teaching of collaborative industrial robots[J]. International Journal of Semantic Computing, 2018,12:167-186.
- [12] TADESEM A, YUMBLAF, PICO N, et al. Application of a reliable dynamic friction model for accurate dynamic model parameters estimation of robot manipulators[C]// 2022 22nd International Conference on Control, Automation and Systems. Jeju, Korea: ICCAS, 2022: 1916-1923.
- [13] 唐辉. 基于机器视觉的咖啡豆智能分级系统开发[D]. 福州: 福建工程学院, 2023.
- [14] 张晋婧,刘双峰,丰雷,等. 融合注意力机制的人脸识别算法研究[J]. 国外电子测量技术, 2023,42(2): 107-113.
- [15] 范雪扬. 基于语音识别技术的在线语言交互学习系统的设计与实现[J]. 自动化与仪器仪表, 2023(7): 187-190,195.
- [16] 梁晴晴. 基于 CNN 与 SVM 的融合特征人脸识别研究 [D]. 石家庄: 河北经贸大学, 2021.
- [17] WEN J. Freshness-aware incentive mechanism for mobile AI-generated content (AIGC) Networks [C] // 2023 IEEE/CIC International Conference on Communications in China. Dalian, China: IEEE, 2023:1-6.

Coffee robot system design with artificial intelligence generated content

CHEN Huashu^{*}, LU Zhenli^{**}, WANG Weidong^{***}

(^{*} School of Mechanical Engineering, Yancheng Institute of Technology, Yancheng 224051)

(^{**} School of Electrical Engineering and Automation, Changshu Institute of Technology, Changshu 215500)

(^{***} East Control Automation Technology Co. Ltd., Changshu 215500)

Abstract

An intelligent coffee robot system is designed based on the East Control's collaborative robotic arm and coffee machine, incorporating the large language generative model of artificial intelligence generated content-generative pre-trained transformer (AIGC-GPT), which is able to achieve voice dialogue based on ChatGPT, and automate the preparation of coffee and coffee pulling based on the voice dialogue. The integration of collaborative robots and GPT through Azure cognitive services (ACS) technology makes the coffee machine no longer just a traditional coffee making device, but a coffee making platform that is intelligent, interactive and personalized. This design includes voice-controlled automated coffee making and ordering, personalized recommendations, and coffee quality control. This innovative convergence is expected to open up new opportunities for the coffee industry and other sectors, increasing user satisfaction and creating a more engaging café experience.

Key words: collaborative robo, ChatGPT, Azure cognitive services (ACS), human-machine interaction