文章编号：2096-1472(2021)-04-55-05
DOI:10.19644/j.cnki.issn2096-1472.2021.04.013

基于云平台的智能语音交互机器人设计

何 松，黄 维，吴昔遥，周曾豪，杨东泽

(空军预警学院，湖北 武汉 430019)
✉339446624@qq.com; 1828824404@qq.com; 1170631815@qq.com;
1364916905@qq.com; 2668430663@qq.com

摘 要：现有的语音交互机器人多采用用户提问、机器人回答的单向交流方式，人机交互的智能性和灵活性较差。本文研究运用树莓派(Raspberry Pi)计算机和配套的语音板作为硬件载体，融合语音唤醒、语音识别、语音合成、自然语言处理等人工智能技术，调用科大讯飞开放云平台、在线图灵机器人，搭建一种基于云平台的智能语音交互机器人系统，并结合自主开发的本地知识库和问题库，使智能语音交互机器人能够根据不同环境与任务需求实现双向互动交流，实现由机器人采集信息和交流反馈，以提供高适应性的无接触人机语音交互服务。

关键词：人工智能；自然语言处理；语音交互机器人；树莓派；云平台
中图分类号：TP24 **文献标识码**：A

Design of Intelligent Voice Interactive Robot based on Cloud Platform

HE Song, HUANG Wei, WU Xiyao, ZHOU Zenghao, YANG Dongze

(*Air Force Early Warning Academy*, *Wuhan* 430019, *China*)
✉339446624@qq.com; 1828824404@qq.com; 1170631815@qq.com;
1364916905@qq.com; 2668430663@qq.com

Abstract: Existing voice interactive robots mostly use user questions and the one-way communication method of robot answers, which is less intelligent and flexible in human-computer interaction. This paper proposes to build an intelligent voice interactive robot system based on cloud platform. The proposed system uses Raspberry Pi computer and the supporting voice board as hardware carriers, and integrates artificial intelligence technologies such as voice wake-up, voice recognition, speech synthesis, natural language processing. It also makes use of the services of IFLYTEK open cloud platform and online Turing robot. Combined with self-developed local knowledge base and question library, the intelligent voice interactive robot can conduct two-way interactive communication according to different environment and task requirements, collect information, and exchange feedback. It provides highly adaptable contactless human-machine voice interaction service.

Keywords: artificial intelligence; natural language processing; voice interactive robot; Raspberry Pi; cloud platform

1 引言(Introduction)

随着人工智能技术的快速发展，深度学习在语音技术领域取得突破性进展[1]。与此同时，在互联网快速发展的驱动下，云端技术架构不断成熟稳定，基于语音的人机交互技术应用越来越广泛，涵盖教育、医疗、家居等各行业领域[2]，如服务机器人、情感交互机器人、教育机器人等[3]。在语音交互方面，云端保存着由海量数据通过深度学习训练而成的各种模型[4]，并通过其强劲的处理能力为终端提供诸如语音识别、语义理解、语音合成等计算量较大的服务[5]。

语音识别技术的研究工作始于20世纪50年代，至今已经走过70年的历程[6]，在国内外被广泛研究[7]。2011年的苹果第四代语音助手Siri的出现，带来了国外语音交互产业的高峰期[8]；2014年亚马逊的智能音箱Echo是人机交互技术进入家庭电器产业的重要节点[9]。随着深度学习算法升级、硬件计算能力提升，大量数据不断参与训练优化模型，语音识别和自然语音理解不断取得突破性进展，国内领先的科大讯飞、百度等公司语音识别准确率达97%以上[10]。语音交互技术链条不断成熟，让机器人具备语音交互功能已然成为一种趋势。本文主要研究

整合现有资源，调整传统语音交互系统软件设计方案，基于云服务平台和ROS(Robot Operating System)框架，设计智能语音交互系统，并且可以安装于小型集成计算机上。作为安装于病房内的“智能语音交互机器人”，降低语音交互系统开发难度和研发成本，使无接触式就医得以实现，并且扩展应用功能。

2 语音交互模块需求分析(Demand analysis of voice interaction)

通过分析疫情防控与病情监测等环境的需要，我们构想将自然语言处理(Natural Language Processing，NLP)技术整合应用至机器人，实现信息采集和交流反馈的主要功能，最终以文本+语音的形式进行输出。

人工信息采集工作重复、枯燥且效率较低，采用机器人进行信息采集相当于机器人提问，人来回答，可以实现自动化、智能化和高效化。信息采集技术路线：(1)预置问题库；(2)将问题文本转语音输出；(3)采集用户回答的语音；(4)调用语音识别模块将语音转文字；(5)提取用户回答的文本中的关键词信息；(6)将对应的问题和回答作为采集的信息存入数据库。

交流反馈则是机器人通过采集声音信号，检测语音信息，传输至本地知识库和云端服务器中寻找相应匹配信息，确认答案后以语音和文本的形式输出反馈。人机交互式的信息采集与交流反馈既可以从病毒传播途径上降低传染风险，又能够利用预设问答库完成反馈，有效节约了人力资源。

3 智能语音交互机器人总体设计(Overall design of intelligent voice interactive robot)

智能语音交互机器人的整体框架有两层：云端服务平台位于云端服务器(本例中采用科大讯飞开放云平台作为云端服务器)，云端保存着由海量数据通过深度学习训练而成的各种模型，可降低终端的解析压力，为系统提供一系列在线支持，主要是对本地计算机向云端发送的数据包提供解析、反馈与下载等服务——包括语音识别、语义理解、语音合成等。本地计算机交互系统主要分为三层：最底层是物理层，为Linux内核，是系统运行环境(本例中采用Ubuntu 16.04系统)，对应的机器人操作系统ROS版本为kinetic；其次是中间层，该层主要是第三方库以及ROS操作系统；最上层称为应用层，主管系统的业务处理逻辑，可根据任务需要设置功能模块区，如在隔离病房中的计算机需具备“病情汇报”“内外交流”等功能。系统架构如图1所示。

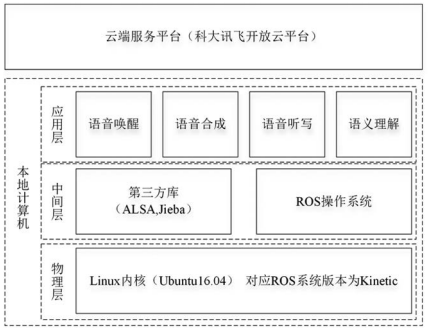

图1 系统架构设计

Fig.1 Design of system architecture

3.1 基于在线云平台的语音交互

智能语音交互机器人主要模块包含语音采集、语音唤醒、语音检测、云端识别、本地知识库检索、图灵机器人交互、语音合成、输出设备播放、判断结束，从而构成逻辑完整、满足功能需要的语音交互系统。在线语音交互流程如图2所示。

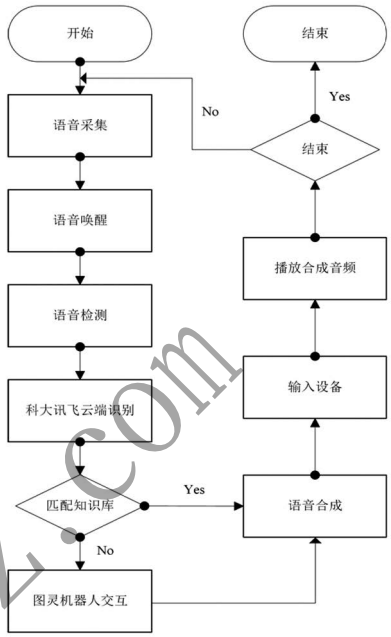

图2 在线语音交互流程图

Fig.2 Flow chart of international voice interaction

语音采集是控制麦克风采集音频将其转换为可供后续流程使用的信号，在系统中以wav文件形式保存。定义get-audio()函数来设定麦克风使用参数，如CHANNELS(声道数)和RATE(采样率)等，从而实现麦克风采集数据的功能。

语音唤醒环节是通过识别输入的音频信号中特定的词语，当识别引擎计算得分超过预设的门限值时，返回唤醒结果为“真”，回调预设的函数，进行下一步的处理。

语音检测是对采集的音频文件进行分析，判断是否有有效语音数据输入，并且检测语音文件是否符合识别要求，实际上是对环境噪声等低相关性的过滤，以及对不规范音频文件的筛选。通过对CHANNELS(声道数)和RATE(采样率)等参数的调用判别，提高采样精度，则能更好完善用户体验。

语音识别模块是将语音输入转为文本输入的过程，基于科大讯飞开放平台所提供的适用于Linux系统下的软件开发工具包(Software Development Kit，SDK)，本地计算机对语音提问进行录制、检测并上传至云端识别引擎，转换成文本数据后，通过互联网重新返回到使用者终端。通过val=os.popen()函数执行讯飞SDK包并将返回结果保存在文件中，并利用函数readlines()循环查询找到语音识别的结果，截取结果，输出到用户终端。因此，到“科大讯飞云端识别”的步骤为止完成了对用户的语音提问转为文本数据的过程，后续步骤会进一步提交文本数据进行问答匹配与语音合成输出。

检索知识库是先读取本地知识库，将语音识别出的文本数据导入其中匹配，若存在匹配项，则返回为“真”，并将

匹配文本数据导入科大讯飞SDK文件，转语音合成输出；若不存在匹配项，则调用图灵机器人，将识别出的文本数据进行在线检索。Fo=open()函数用来打开知识库，readlines()函数将知识库读取为列表变量，进行结果查询。本步骤完成了问答匹配流程，并提供了本地知识库匹配和云端知识库匹配两种途径。

语音合成即从文本输入到语音输出的过程，把知识库匹配的答案上传至云端服务器，转为语音数据后返回用户终端播放。在线语音听写和在线语音合成都属于通过音频文件/文本文件向云端识别引擎请求服务并获得识别结果的方式，相较于建立传输控制协议/互联协议(Transmission Control Protocol/Internet Protocol，TCP/IP)长连接的方式发送实时音频数据流的方式，前者实时性较差但不必长时间占用计算机资源。通过os.popen()函数执行科大讯飞的语音合成SDK文件，以匹配所得的文本数据为对象，生成wav格式的音频文件，并将该文本数据输出到终端，实现语音和文字两种形式输出。

设备输出音频是通过调用生成的语音文件输出给外设麦克风实现的。利用os.popen()函数调用play指令播放音频，并根据合成音频文件的文本数据长度决定暂停时间长短，保证语音输出的完整性。

3.2　离线语音交互

针对涉密或隐私情况下离线语音交互的需要，可将基于云端服务器的识别处理转为发送至本地计算机进行。通过更改各节点间的订阅关系，将原发送至云平台的数据转发给本地程序处理，以实现离线的语音交互。

其基本步骤是：语音交互系统启动后，由用户输入指定唤醒词，将系统由待机状态唤醒至工作状态，调用system()函数对用户的提问进行录音，生成指定wav文件。利用回调函数将该wav文件输入科大讯飞离线语音识别SDK进行识别，识别理解转化成文本数据返回。调用system()函数对文本数据进行获取，将问题文本数据导入预先编好的本地语料库中进行匹配，得到对应的回复文本数据。利用回调函数将该回复文本数据输入科大讯飞离线语音合成SDK进行合成，得到特定内容的wav音频文件。最后调用system()函数对合成的wav音频文件进行播放，即完成一次完整的语音交互过程。

3.3　基于语音交互的信息采集

信息采集功能是通过机器人主导交互实现的。利用科大讯飞离线的语音合成SDK文件将设定的问题处理为语音，再通过扬声器播放出来，被采集者听到问题的反馈将被麦克风收录，SDK文件将语音文件识别为文本，提取关键词。作为采集信息，机器人将问题和对应采集信息存入数据库，实现了无接触的信息采集功能。

交流反馈功能是通过使用者主导交互实现的。先通过语音唤醒，使机器人调用麦克风采集声音，然后将声音信号发至基于科大讯飞开放平台的语音识别模块，提取识别结果并在知识库中检索，将得到的相关文本发至语音合成模块，再将输出结果连入扬声器播放，最终反馈至使用者。信息采集与交流反馈流程图如图3所示。

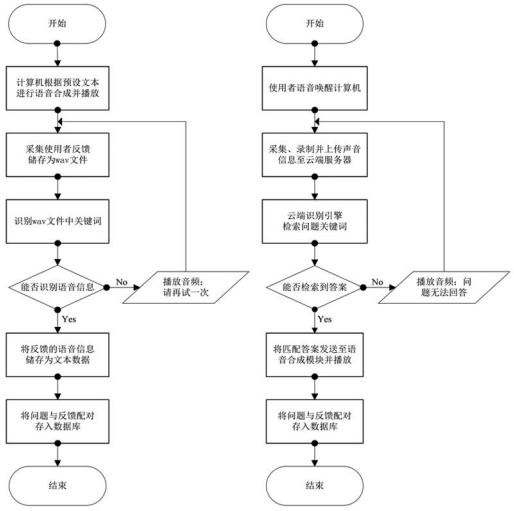

图3 信息采集与交流反馈流程图

Fig.3 Flow chart of information collection and communication

3.4　语音交互机器人在病房中的实际应用

经过对体积与应用性的考量，该机器人采用树莓派(M4PB型)作为硬件载体，以体积小和集成度高适应便携性需求，并且应用了科大讯飞开放云平台的在线资源和本地语料库，通过将语音数据经由网络传输到云端服务器，利用云计算技术得到识别结果并返回。云端资源丰富，可以更好地利用知识库资源，同时本地语料库的准备适用于不同情况下的不同需要，分析设计其相对应的资料库，并延伸相关的可靠性设计与适应性设计，可以实现更广泛的用途。语音交互系统采用ROS节点消息发布和订阅机制。ROS是开源的机器人操作系统软件，提供类似于操作系统的服务。ROS通过将庞大繁杂的系统任务切分成功能单一的子任务，再通过以消息或服务的方式将子任务链接起来形成可以完成复杂任务的系统，实现代码复用，降低设计难度，同时ROS支持C、Python多编程语言，功能包丰富，测试方便。在通信过程中，节点将消息以特定主题发布到ROS核心控制器，ROS核心控制器异步地将该消息转发给订阅该主题的节点进而实现通信。ROS节点消息传递示意图如图4所示。

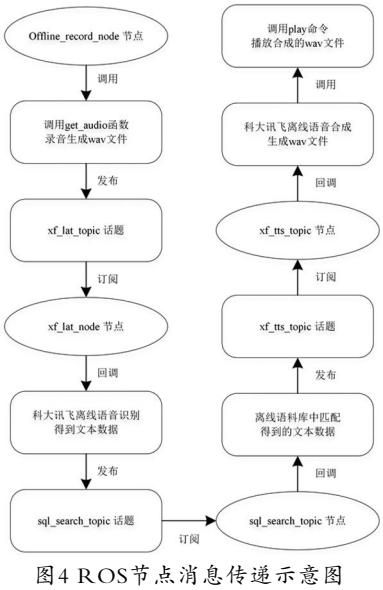

图4 ROS节点消息传递示意图

Fig.4 Sketch map of ROS node message passing

4 优化设计(Optimal design)

由于采样识别的开放式环境会带来大量数据和样本，易造成数据冗杂和过拟合问题，并且计算机一一识别将会提高识别难度并增加运算时间，进而降低产品的使用寿命，因此语音数据预处理和特征工程，即对数据进行降噪、转换和分类的专项筛选处理可以节省大量资源并提高语音交互模型性能。本系统从以下四个方面进行处理。

(1)特征提取

特征提取是通过以相对较低的数据采样速率将波形数据转换为参数表示形式，而后进行后续处理和分析来完成的。这通常称为前端信号处理，它将经过处理的波形语音信号通过函数如感知线性预测(PLP)、线性预测编码(PC)和频率倒谱系数(MFCC)，转换成一种简洁而有逻辑的表示形式，比实际信号更有鉴别性和可靠性。

(2)特征降维

数据降维是在降低特征数量的同时，尽可能保留原数据主要的信息，利用同一特性的最优特征筛除冗余特征，最终得到对构建模型最有贡献度的特征。降维处理后的数据集合具有更小的规模，这样的集合更易于储存并且可以有效降低运算的复杂性，还可以大幅降低模型的复杂性，防止过拟合的情况出现。

(3)特征过滤

特征过滤是特征选用方法中最为常见和基本的一种，可以通过设立阈值来限制无关数据的输入，比如在唤醒程序中设置音量阈值，可以简单有效地过滤外部杂音和自身移动碰撞产生的噪音，大幅降低运算的复杂程度和运算资源的占用量。

(4)特征构造

特征构造是建立使用者或使用环境下常见有效输入的声学模型(例如在病房中使用时构造出现频度高的医学词汇)，通过近似模型来过滤差异较大的无关信息，将拟合程度高的特征信息输入系统，大幅减少数据处理量。

语音材料的预处理在特征工程之前进行。预处理的步骤是预强调(滤波)—帧阻塞(将语音信号按帧分割)—语音信号加窗(加汉明窗和矩形窗对信号进行均匀化处理)，以及必要的降噪与放大处理等。

5 实验结果与分析(The results and analysis of experiment)

5.1 实验环境

硬件配置由树莓派4 B、树莓派3 B+和语音版组成，内置4核处理器ARMv7 process rev3，主频1500 MHz，内存容量1.00 GB，磁盘容量16 GB。搭载Ubuntu 16.04 LST系统+ROS系统，Linux内核版本为4.19.75-v71，ROS系统版本为kinetic。实验环境配置如表1所示。

表1 实验环境配置

Tab.1 Experiment configuration states

名称	树莓派4 B	树莓派3 B+
CPU	1.5 GHz，Quad-Core Broadcom BCM2711B0 (Cortex A-72)	1.4 GHz，Quad-Core Broadcom BCM2837B0 (Cortex A-53)
内存	1—4 GB DDR4	1 GB DDR2
GPU	500 MHz VideoCore5	400 MHz VideoCore3
视频输出	双微型HDMI接口	单HDMI接口
最大分辨率	4 k 60帧+1080 P或 2×4 k 30帧	2560×1060
USB接口	双USB 3.0+双USB 2.0	四USB 2.0
网络连接	千兆以太网	330 Mbps以太网
无线连接	802.11ac(2.4/5 GHz)，蓝牙5.0	802.11ac(2.4/5 GHz)，蓝牙4.1
尺寸	88 mm×58 mm×19.5 mm	82 mm×56 mm×19.5 mm
重量	46 g	50 g

5.2 实验结果

在线语音交互流程是：(1)人与智能机器人进行语音交互；(2)智能机器人通过麦克风对交互语音进行采集，生成语音wave文件；(3)语音识别节点通过互联网将wave语音文件传输到科大讯飞语音识别服务器，科大讯飞语音识别服务器通过智能语音识别算法将语音文件识别并转换成文本文件，通过互联网发回智能机器人终端；(4)语言处理节点将识别出的文本通过互联网发送到在线图灵机器人；(5)在线图灵机器人通过传入的文本内容和前后文语境，在知识库中查找最佳的回复信息，并通过互联网传回智能机器人终端；(6)语音合成节点收到图灵机器人的文本回复信息后，将其再次发送到科大讯飞云在线语音合成平台；(7)科大讯飞云在线语音合成系统将文本内容转换成语音数据，以MP3格式文件发给智能机器人；(8)智能机器人通过音频输出接口播放回复的语音文件，完成语音数据输出。询问天气的语音交互过程如图5所示。

图5 在线语音交互实验结果

Fig.5 The result of voice interactive test

基于语音交互的信息采集是机器人提问，由人来回答问题，完成信息采集任务。可将问答文本实时合成语音，也可提前把设定好的信息采集音频合成好，不用每次信息采集的时候当场合成，提高程序运行效率。由机器人询问并采集某用户姓名、编号、体温的语音交互过程如图6至图8所示。

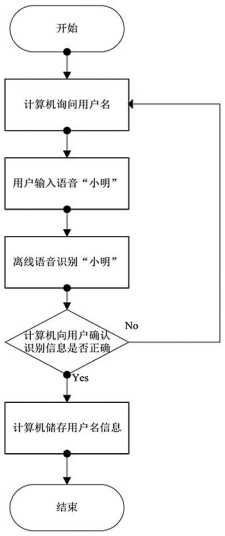

图6 机器人采集用户姓名的语音交互过程

Fig.6 Flow chart of user's name collection

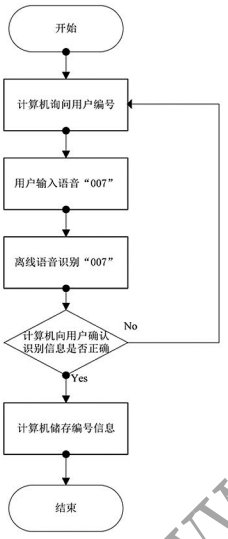

图7 机器人采集用户编号的语音交互过程

Fig.7 Flow chart of user's number collection

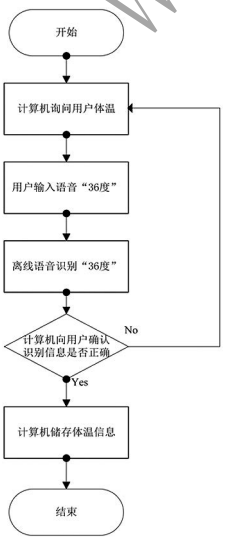

图8 机器人采集用户体温的语音交互过程

Fig.8 Flow chart of user's body temperature collection

通过测试，得到语音交互系统的各项参数如表2所示。

表2 语音交互系统参数

Tab.2 System parameters of Voice Interaction

参数	数据
语音识别准确率	>96%
问题反馈准确率	>79.6%
响应时间	<5 s

6 结论(Conclusion)

在人工智能技术飞速发展的今天，智能化的设备已经融入人们生活的方方面面，提高了生活的便捷性。本文设计了应用于疫情防控与病情监测的ROS智能语音交互机器人，通过对录入音频信号的预处理和特征提取，利用科大讯飞SDK文件和图灵机器人模块，以及有针对性的知识库，同时充分考虑信号复杂性，设计降噪滤波方案，实现了无接触式智能语音交互，减轻了医护人员的工作量并从传播途径上降低了感染风险。由于并未考虑多阶段对话中复杂逻辑交互的情况，对话时逻辑复杂会对语义理解造成不利影响，比如上下文理解困难、微型机算力不足等，因此设计并优化多阶段复杂逻辑的识别和处理能力将是下一步研究的重点。

参考文献(References)

[1] 戴礼荣,张仕良.深度语音信号与信息处理:研究进展与展望[J].数据采集与处理,2014,29(02):171-179.

[2] 林枫亭,罗艺,孔凡立,等.一种基于云平台的智能机器人语音交互系统设计[J].电子测试,2018(Z1):40-42.

[3] 杨国庆,黄锐,李健,等.智能服务机器人语音交互的设计与实现[J].科技视界,2020(09):129-131.

[4] 秦伟.基于语音的人机交互平台的设计与实现[D].武汉:华中科技大学,2019.

[5] Shenzhen Aukey Smart Information Technology Co., Ltd.. "AI Voice Interaction Method, Device And System" in Patent Application Approval Process (USPTO 20200105268)[J]. Telecommunications Weekly, 2020.

[6] YAO D, KATIE S T. Bridging the gap in mobile interaction design for children with disabilities: Perspectives from a pediatric speech language pathologist[J]. International Journal of Child-Computer Interaction, 2020:23-24.

[7] 杨加平.面向指控系统的嵌入式语音交互技术设计与实现[J].机械与电子,2015(04):72-74.

[8] 廖彬全,罗佩,马远佳.基于智能语音交互系统的翻译机器人[J].信息与电脑(理论版),2019,31(17):110-112.

[9] 陈鑫源.智能语音交互技术及其标准化[J].电声技术,2018,42(05):78-80.

[10] 郝欧亚,吴璇,刘荣凯.智能语音识别技术的发展现状与应用前景[J].电声技术,2020,44(03):24-26.

作者简介：

何　松(1988-)，男，硕士，讲师.研究领域：模式识别，人工智能.

黄　维(1999-)，男，本科生.研究领域：自然语言处理.

吴昔遥(1999-)，男，本科生.研究领域：自然语言处理.

周曾豪(2000-)，男，本科生.研究领域：自然语言处理.

杨东泽(2000-)，男，本科生.研究领域：自然语言处理.