

城市轨道交通智能语音售票机关键技术及应用研究

陈悦勤 钱小毅

(上海申通地铁集团有限公司, 201103, 上海//第一作者, 工程师)

摘要 智能语音售票机具有语音识别功能,乘客通过语音可完成购票流程,但地铁车站环境复杂、背景声音嘈杂、人流密集,传统的语音识别技术难以消除噪声,因而难以使乘客获得良好的语音购票体验。分析了新型的智能语音售票机系统的麦克风阵列语音增强技术、人脸关键点识别技术和多模态信息融合技术,介绍了该系统的硬件架构和购票业务流程。应用效果表明,该系统可有效提高语音信噪比,可满足车站复杂环境下高准确率、低时延的语音处理需求,可有效改善乘客语音购票体验,能够缩短乘客平均购票时间。

关键词 城市轨道交通;售检票系统;智能;语音识别

中图分类号 U293.22

DOI:10.16037/j.1007-869x.2022.10.028

Research on Key Technology and Application of Intelligent Voice TVM to Urban Rail Transit

CHEN Yueqin, QIAN Xiaoyi

Abstract The intelligent voice TVM (ticket vending machine) has the function of voice recognition, and passengers can complete the ticket purchase process through voice. However, the metro station features complex environment, noisy background sound and dense crowd, the conventional voice recognition technology has difficulty in eliminating the noise, holding back a good voice ticket purchase experience for passengers. The microphone array speech enhancement technology, face key point recognition technology and multi-modal information fusion technology of the new intelligent voice TVM are analyzed. The hardware architecture and ticket purchase process of the system are introduced. Application results show that the system can effectively improve the voice signal-to-noise ratio, meet the voice processing requirements of high accuracy and low delay in the complex environment of the station, effectively improve the voice ticket purchase experience of passengers, and finally shorten the average ticket buying time of passengers.

Key words urban rail transit; fare collection system; intelligence; voice recognition

Author's address Shanghai Shentong Metro Group Co.,

Ltd., 201103, Shanghai, China

0 引言

发展城市轨道交通是建设绿色城市、智慧城市的有效途径。智慧车站基于全面的智能感知、深度的数据融合、高效的数据治理理念,利用多种智能化手段,实现丰富乘客乘车体验、提升乘客满意度、提高设备设施可靠性及实用性等目标^[1]。

简便购票流程、提高购票效率能够有效提升乘客满意度。传统的乘客购票方式主要有窗口购票和自动售票机(TVM)购票两种。自动售票机能够快速实现线路查询、车票发售、找零等功能,相较于窗口购票,其能够有效提高乘客的购票效率,节约乘客购票时间。但当乘客不熟悉操作流程或不识字时,可能造成购票耗时的增加,甚至可能因操作不当引起设备故障以及增加窗口的人工负担^[2]。

智能语音售票机引入了语音识别功能,能够帮助不熟悉的乘客完成购票全部流程,以避免发生上述问题。乘客购票时,可通过触发语音切换按钮将手动操作转化为单纯语音输入操作,高效便捷地实现票价咨询、站点信息查询、换乘信息查询、嘈杂环境下语音正确识别、车票购买等功能^[3]。仅需通过口述目的站点及车票张数等信息即可实现正确购票,大大简化了购票流程,提高了购票效率,也减轻了票务人员人工操作压力。

智能语音售票机需具备嘈杂环境语音识别能力。公众空间混合强噪声环境中的多人人声、车辆行驶轰鸣声、广播播报声、大封闭空间产生的混响等多种因素会严重影响语音识别的效果,因此该技术的关键及难点在于噪声处理^[4]。此外,在该环境下售票机的唤醒方式不宜采用语音唤醒,人脸识别技术能够更加有效地保证设备唤醒效率^[5]。

目前,智能语音辅助购票技术已应用于上海轨

道交通部分车站的自动售票机,在乘客购票的关键步骤中提供“因人而宜”的个性化语音提示,使乘客在购票准确度、购票服务支持等方面的体验有了较为明显的提高^[6]。

本文针对地铁车站嘈杂环境和实际用户需求,探讨新型智能语音售票机系统的线性麦克风阵列语音增强技术、基于视频的人脸识别技术和多模态信息融合技术应用,该系统可实现售票机“免唤醒”,以及高效的智能语音购票和地点查询功能,可有效提升乘客购票效率。

1 智能语音售票机关键技术

为了应对嘈杂环境对智能售票机语音识别能力的影响,智能语音售票机系统采用了线性麦克风阵列语音增强技术、基于视频的人脸识别技术和多模态信息融合技术。

1.1 麦克风阵列语音增强技术

图1为麦克风阵列拾音示意图。

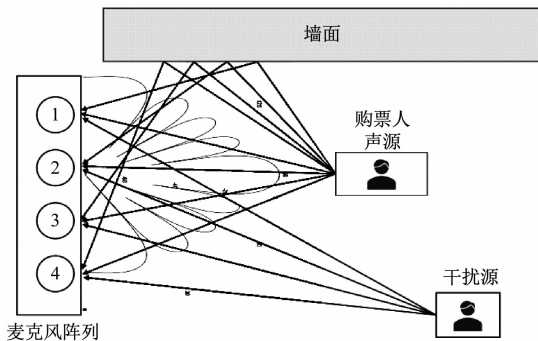


图1 麦克风阵列拾音示意图

Fig. 1 Schematic diagram of microphone array pickup

常见的单麦克风前端语音增强方法虽然能实现大量降噪,但对非平稳噪声抑制能力较弱,同时会使引入的语音信号失真,导致在地铁车站等噪声动态范围大、噪声模式多样的场景下难以获得较高的语音信号信噪比,无法满足语音识别的要求。

均匀线性阵列示意图如图2所示。图2中, M 个麦克风以 δ 为间隔,接收由声源发出的声音信号 $X(f)$ (f 为频率),声波入射角度为 θ_d ,第 i 个麦克风接收信号为 $Y_i(f)$, $V_i(f)$ 为第 i 个麦克风的加性高斯白噪声。

麦克风接受信号组成的矢量 $\mathbf{Y}(f)$ 为:

$$\mathbf{Y}(f) = \mathbf{d}(f, \cos \theta_d) \mathbf{X}(f) + \mathbf{V}(f) \quad (1)$$

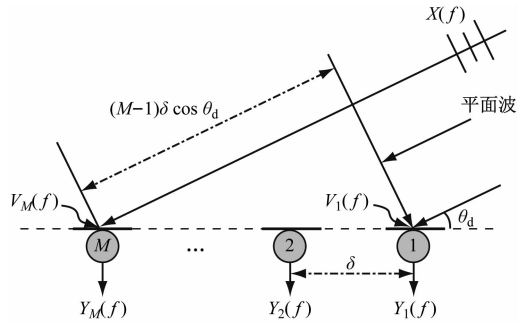


图2 均匀线性阵列示意图

Fig. 2 Schematic diagram of uniform linear array

式中:

$\mathbf{d}(f, \cos \theta_d)$ ——阵列的导向矢量;

$\mathbf{X}(f)$ ——由声源发出的声音信号矢量;

$\mathbf{V}(f)$ ——麦克风的加性高斯白噪声矢量。

$$\mathbf{d}(f, \cos \theta_d) = [1 \quad e^{-j2\pi f \tau_0 \cos \theta_d} \quad \dots \quad e^{-j(M-1)2\pi f \tau_0 \cos \theta_d}]^T \quad (2)$$

式中:

j ——虚数;

τ_0 ——声波入射角度 $\theta_d = 0$ 时相邻两个麦克风之间的时间延迟。

麦克风阵列的主要性能指标包括波束图和白噪声增益。波束图可表述阵列在特定频率 f 对不同入射角信号的增益 $B(f, \theta_d)$,应用延迟求和波束成形算法的波束增益公式为:

$$B(f, \theta_d) = \mathbf{d}^H(f, \cos \theta_d) \mathbf{h}(f) \quad (3)$$

其中: $\mathbf{h}(f) = [H_1(f) \quad H_2(f) \quad \dots \quad H_M(f)]$, $H_i(f)$ 表示第 i 个麦克风在频率 f 上的响应; \mathbf{d}^H 表示对向量 \mathbf{d} 作共轭转置。

白噪声增益为空间非相干白噪声场下的阵列增益,即指向方向的增益与平均噪声功率之比。当采用延迟求和算法时,其阵列增益 $W(f)$ 表达式为:

$$W(f) = M \quad (4)$$

可见,白噪声增益与麦克风数量 M 成正比,麦克风数量越多,指向方向获得越好的抗噪性能。

1.2 基于视频的人脸识别技术

地铁环境声音嘈杂,也无法对乘客提前进行语音购票机使用培训,采用“唤醒词+命令句式”实现语音购票功能唤醒不适用于地铁车站环境。智能语音售票机系统采用视频摄像头检测人脸,确定人员正在购票,从而唤醒语音购票功能。

地铁车站环境视频检测复杂性较高,存在大量背景杂物,人脸姿态角度不一,光照和遮挡也会对

检测造成影响,进行快速准确的人脸检测难度较大。

本文基于多任务卷积神经网络(MTCNN)架构设计了一种人脸匹配算法,对人脸特征进行多重检测后提取人脸信息。其网络构架(见图3)主要包含 PNet(Proposal Network, 候选网络)、RNet(Refine Network, 修正网络)和 ONet(Output Network, 输出网络),分别负责人脸的初步检测、位置修正、复杂样本分类,最终输出人脸位置和关键点信息。

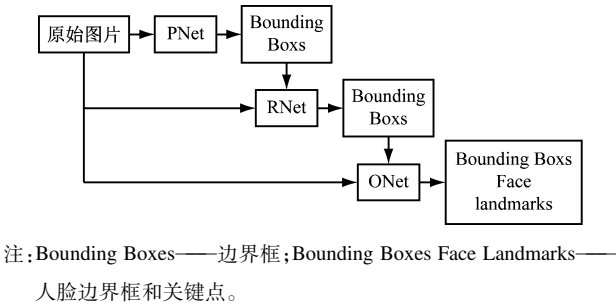


图3 MTCNN 网络结构

Fig.3 MTCNN network structure

1.3 多模态信息融合技术

多模态信息融合是指整合或融合两种及两种以上信息来源,利用不同信息媒介的优势,使得识别或信息处理更加精准。在智能语音售票机系统中,多模态体现在以下两个方面:

1) 音视频协同处理:可同时采集高清视频流、麦克风阵列信号和用户操作信息,通过用户点击“语音购票”按钮或摄像头检测到有人走到售票机屏幕前,语音模块自动唤起。其采用免唤醒语音输入,不需要输入其他同类产品的唤醒词,免去了用户培训过程。此后,摄像头通过用户脸部关键点来定位用户声源位置,麦克风阵列通过8个麦克风孔来进行波束成形,将主瓣对准用户,结合摄像头和麦克风阵列来判断当前说话用户的位置是否一致,一致则用户语音有效,不一致则用户语音无效。即通过视频识别与声音识别进行多模态信息融合判断。

2) 地图知识库支撑的地铁站点查询:一般语义识别方案仅支持查询特定地铁站名作为购票起止点,不具备模糊查询功能,如根据地标设定附近车站。智能语音售票机系统采用外部地图的成熟技术数据,协同语义识别的起止点信息,自动反推最近的地铁站,大大提高了语音售票机服务的便捷性和易用性,更适应乘客常用的起止点表达方式。

2 智能语音售票机系统

基于上文所述的麦克风阵列语音增强技术、人脸检测算法和多模态信息融合技术,在传统 TVM 基础上增加了智能语音模块,使其具备车站复杂环境下的语音购票功能。本节重点介绍具体硬件架构和购票业务流程。

2.1 硬件架构

智能语音售票机系统硬件架构(见图4)包含以下主要模块:

- 1) TVM 主控:地铁站现有的售票机箱体机器通过串口与智能处理工控机进行数据交互。
- 2) 智能处理工控机:负责麦克风阵列语音增强处理和视频人脸检测处理。
- 3) 语音处理云服务器:部署在地铁运营公司数据中心,是一台专有云语音识别和自然语言理解服务器,由地铁专网保障其网络信息安全。
- 4) 第三方地图信息服务器:部署在地铁运营公司数据中心,是一台专有云第三方地图服务器,为语音购票技术提供地理位置查询功能。
- 5) 高清摄像头和麦克风阵列:工业级别的摄像头和麦克风阵列,需要在 TVM 机器上打洞安装,通过 USB 与信号处理工控机通信。
- 6) 数据回传模块:由无线或有线网络连接至 AFC(自动售检票)专网,将语音和视频信息经过数据安全性增强后接入专有云服务器。

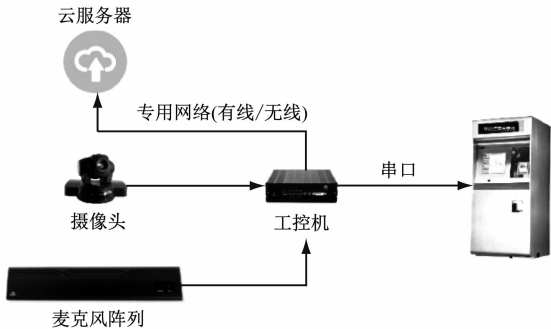


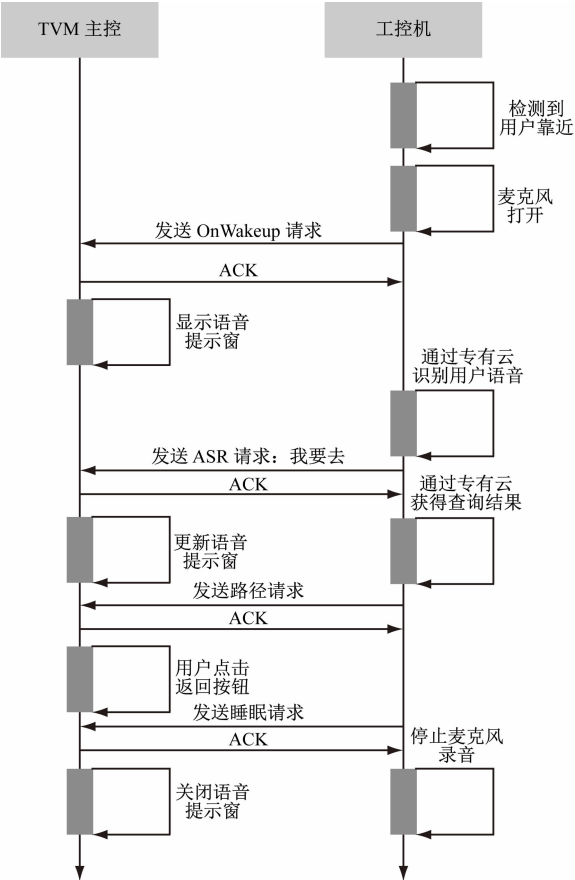
图4 智能语音售票机硬件架构

Fig.4 Hardware architecture of intelligent voice TVM

2.2 典型交互流程

语音购票过程中的典型人机交互流程如图5所示。

当工控机通过高清摄像头检测到用户靠近时,会主动发送 OnWakeup(唤醒指令)请求给 TVM 主控。TVM 主控需返回确认唤醒 ACK(确认唤醒),



注：ASR 为语音业务请求。

图 5 语音购票典型人机交互流程

Fig. 5 Typical human-computer interaction process of voice ticket purchase

并显示语音提示窗口。

随着用户说话的过程,工控机会多次发送 ASR,逐字返回用户语音内容。TVM 主控需要多次刷新语音提示窗口中的语音内容。当用户说完以后,工控机查询到结果,则会发送路径请求给 TVM 主控,直接跳转路径导航页面。如果有超过一个候选项,则显示多个候选项。

在语音提示窗口状态下,用户点击屏幕其他区域或者点击返回按钮,则 TVM 主控发送睡眠请求,并关闭语音提示窗口。

2.3 主要功能

1) 语音问询:乘客在问询时,可以直接通过语音说出要去的地点,问询机会推荐附近的地铁站和引导相关购票流程。

2) 免唤醒启动:监测到用户走到问询机器前时,语音模块会自动唤起,随即进入问询初始化界面,乘客从而可以直接进行语音问询。

3) 地点确认:用户每说完一句话,经过语音转

文本、语义分析后,结合第三方地图信息查询到最近站点,给出出行建议。如果有存在多个地点或者多个地铁站的情况,将备选项(包含地点、附近地铁站推荐)显示在界面上供用户选择,最多显示 5 个。

3 应用效果

智能语音售票机需具有良好的实时响应性能和准确的乘客语音识别能力。因此,需能够实时监测乘客靠近、高抗扰语音识别和提供无延迟综合服务。

本文从麦克风阵列的抗噪性能、智能 TVM 总体技术性能和用户体验 3 个角度分析应用效果。

3.1 麦克风阵列语音抗噪性能

通过仿真计算得到基于 8 个麦克风的线性阵列和语音增强技术的波束图,如图 6~7 所示。图 6 中,两侧旁瓣相对于正前方主瓣至少弱 18 dB,图 7 中的也至少弱 13 dB。可见,该语音增强技术可有效提升正前方声音增益,屏蔽侧方其他 TVM 的乘客语音。结合人脸识别和关键点定位,可进一步在时域或空间域处理目标语音,消除其他方向的干扰。

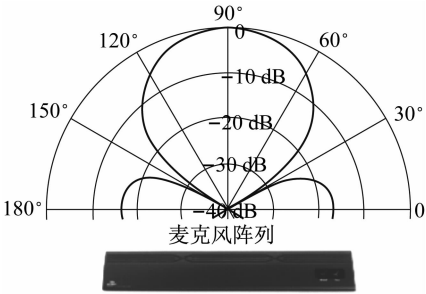


图 6 麦克风阵列波束图(频率 1 000 Hz)

Fig. 6 Microphone array beam diagram (frequency of 1 000 Hz)

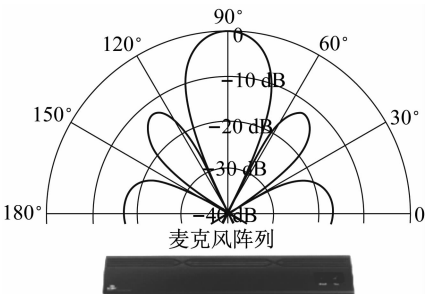


图 7 麦克风阵列波束图(频率 2 000 Hz)

Fig. 7 Microphone array beam diagram (frequency of 2 000 Hz)

3.2 TVM 总体技术性能

在实际应用中,通过测试验证,该智能语音售

票机可满足以下性能指标:

- 1) 在 80 ~ 90 dB 的强噪声环境下仍然可以正常工作;
- 2) 满足人距离设备 1 m 以内的语音交互场景;
- 3) 支持免唤醒语音输入、拼音首字母输入和模糊地点查询功能;
- 4) 设置地铁车站名词 415 个(涵盖所有地铁站名),导入上海地标、景区、地铁周边设施、小区等名词近 3 000 个(几乎涵盖所有地点);
- 5) 地址名词正确率 > 90% ;
- 6) 网络通畅时,时延 ≤ 0.5 s;
- 7) 支持地址模糊查询;
- 8) 满足 IP54(防尘防水等级)、部署、长期维护和检测、数据统计等各类其他要求。

3.3 用户体验

在上海轨道交通汉中路站进行了传统售票机和智能语音售票机的购票测试。共采集两种售票机的购票时间数据样本 89 个。购票时间是指用户开始与 TVM 交互,至购票结束为止所用时间。传统售票机用户平均购票时间为 30.64 s,而智能语音售票机的仅需 11.85 s,大大提升了售票效率。

4 结语

本文探讨了智能语音售票机系统麦克风阵列语音增强技术、人脸关键点识别技术及多模态信息融合技术,介绍了具体硬件架构和业务流程。智能语音售票机系统可实现乘客智能语音购票、缩短乘客购票时间、减少购票排队压力,同时也能有效降低设备故障率及维修人力、物力的投入,提升服务

质量。

参考文献

- [1] 林晓伟.城市轨道交通智慧车站设计与实现[J].工业控制计算机,2020(11):109.
LIN Xiaowei. Design and implementation of urban rail transit smart station[J]. Industrial Control Computer, 2020(11):109.
- [2] 郭兴宝.关于智慧车站城市轨道交通中发展的前景分析[J].中小企业管理与科技(中旬刊),2020(7):124.
GUO Xingbao. Analysis of the development prospect of smart station in urban rail transit[J]. Management & Technology of SME, 2020(7):124.
- [3] 陈韵舟.轨道交通智慧车站研究[J].科技创新与应用,2020(16):62.
CHEN Yunzhou. Research on smart station of rail transit[J]. Technology Innovation and Application, 2020(16):62.
- [4] 焦科杰,刘清乐.语音识别技术在轨道交通智能票务终端中的研究与应用[J].通信与广播电视,2018(增刊1):30.
JIAO Kejie, LIU Qingle. Research and application of voice recognition technology in rail transit intelligent ticketing terminal[J]. Communication & Audio and Video, 2018(S1):30.
- [5] 曾家建.基于深度学习的人脸关键点检测和人脸属性分析[D].杭州:浙江大学,2018.
ZENG Jiajian. Facial landmark detection and face attribute analysis based on deep learning [D]. Hangzhou: Zhejiang University, 2018.
- [6] 吴明欣.自动售票机智能语音辅助购票系统开发与研制[J].城市建设理论研究(电子版),2013(9):1.
WU Mingxin. Development and research of intelligent voice assisted ticket purchasing system for automatic ticket vending machine [J]. Theoretical Research in Urban Construction, 2013(9):1.

(收稿日期:2021-04-01)

8 月份全国城市轨道交通客运量环比增长 0.33%

截至 2022 年 8 月,31 个省(自治区、直辖市)和新疆生产建设兵团共有 51 个城市开通运营城市轨道交通线路 278 条,运营里程达 9 098 km,实际开行列车 297 万列次,完成客运量 19.6 亿人次。8 月份,客运量环比增加 0.1 亿人次,增长 0.33%,同比增加 2.0 亿人次,增长 11.37%。本月新增运营里程 31 km,新增运营线路 1 条,为福州地铁 6 号线。

按照疫情防控要求,石家庄、乌鲁木齐、大连的地铁和三亚市的有轨电车暂停运营,天津、沈阳、哈尔滨、上海、南京、无锡、温州、郑州、武汉、广州、深圳、重庆、成都等城市有部分地铁车站或车站出入口暂停运营。截至 8 月底,石家庄、乌鲁木齐、大连的地铁和三亚市的有轨电车未恢复运营。

(摘编自 2022 年 9 月 7 日“交通运输部微信公众号”)