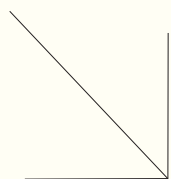


客服调控系统中 多模态交互技术的应用探索

文◆国家电网公司客户服务中心 刘爱玲 居 强 张知宇 程天阳



在此背景下，多模态交互技术成为优化客服调控系统的重要方向。

1 多模态交互技术概述

多模态交互技术是一项集成了多种人工智能技术的创新技术，其通过整合计算机视觉、语音识别、声音克隆和自然语言处理等技术，使人们能够通过声音、肢体语言、文字、图片、音频和视频等多种途径与计算机系统进行互动，从而为用户提供一种更加自然和人性化、沉浸式的交互体验。

常用的多模态交互方式如下。（1）语音与文本交互。用户通过语音输入，系统将语音转换为文本后进行处理；通过文本输入，系统以语音形式反馈。这种方式广泛应用于语音助手和智能客服系统中。（2）视觉与手势交互。系统利用摄像头捕捉用户的面部表情、眼动和手势等信息，并据此作出响应。例如，通过手势控制设备或根据面部表情调整交互内容。（3）触觉反馈。运用触觉设备，用户可感受来自系统的物理反馈，如振动或力反馈，增强交互的沉浸感^[1]。

引言

客服调控系统作为国家电网的重要客户服务平台，承担着海量的用户咨询、业务办理以及紧急调度任务。近年来，随着客户服务需求的多样化，传统基于单模态交互的客服系统难以满足高效精准服务的需求。语音交互虽然提升了信息输入的便捷性，但在应对复杂业务时，单一的语音或文本模式往往导致信息丢失或误判。此外，部分业务涉及故障诊断、图像验证等环节，仅依赖语音或文本难以实现精准匹配。

2 客服调控系统中多模态交互技术的具体应用

多模态交互技术在客服调控系统中实现了信息感知的全维覆盖，既包括语音、文本、图像等多种数据，也涵盖用户情感与行为信号^[2]。

2.1 语音、文本、图像融合的智能问答系统

客服调控系统在多模态交互中整合语音、文本以及图像数据，构建智能问答系统，以应对复杂的电力业务咨询需求。用户在拨打客服热线时，系统首先利用自动语音识别（ASR）技术，将语音输入转化为文本，并结合自然语言理解（NLU）模型解析意图，提取关键词，如“电费查询”“停电报修”等。若用户使用在线文本交互，如移动端或微信公众号，系统直接解析文本输入，并匹配电力知识库提供标准化答案。对于涉及图片信息的咨询，如电费账单解析，用户可上传账单截图，系统利用光学字符识别（OCR）技术提取账单核心数据，包括用户编号、缴费金额以及账单周期，并与电力数据库进行匹配，自动返回对应账单

【作者简介】刘爱玲（1990—），女，山东德州人，硕士，中级经济师，研究方向：电力客户服务。

详情^[3]。当用户咨询设备故障或线路问题时，系统支持上传设备照片，计算机视觉模块结合电力设备知识图谱进行图像分析，识别设备型号以及潜在故障特征，并反馈可能的故障类型与处理建议。智能问答系统还采用多模态融合技术，对用户输入的语音、文本和图像信息进行交叉验证。例如，若用户在语音或文本中提及“电表异常”，同时上传电表照片，系统会比对语音内容与图像解析结果，确保问题描述的准确性。此机制减少了用户在复杂业务办理中的重复输入，并提高了问题匹配的精准度。

2.2 远程故障排查中的多模态诊断机制

客服调控系统在电力故障排查中引入了多模态交互模式，整合语音、视频、图像以及文本数据，实现远程故障诊断。当用户报告用电异常时，系统首先通过语音识别提取核心诉求，如“家中停电”“变压器过载”等，并结合用户历史报修记录进行智能匹配，判断是否为重复故障。若用户选择图片或视频提交模式，系统会引导用户上传电表、电箱或电力设备的现场照片，并利用目标检测算法分析电表读数、设备连接状态以及电缆破损情况。在实时诊断过程中，系统利用增强现实和计算机视觉技术，为用户提供交互式远程指导。客服人员可在视频通话界面中标注关键操作点，如指示用户检查电表开关、测试电流电压或确认是否存在外部线路损坏。对于疑似设备故障的情况，系统调用深度学习模型，对视频画面进行异常检测，识别是否存在短路、烧损或过载现象，并匹配历史维修案例，为用户提供处理建议。若问题无法远程解决，系统会自动生成结构化工单，并附带语音记录、图像分析结果以及视频通话摘要，以便维修人员在现场检修前即可获取完整故障信息。

2.3 智能工单调度中的多模态数据整合

客服调控系统在工单管理环节集成多模态数据分析，以优化任务分配与资源调度。系统在接收客户诉求后，先基于语音、文本以及图像数据进行问题分类，并结合历史数据分析用户过往维修记录，以判断故障模式。当用户多次报告电表故障，系统会综合评估设备老化风险，并优先将工单分配至经验丰富的维修人员。在调度过程中，系统结合地理信息系统技术，获取用户所在区域的电网状态，并结合语音分析、视频上传以及图像解析结果，判断维修需求的紧急程度。若系统检测到“电力中断”且用户上传的照片显示设备损坏，则自动提升工单优先级，并结合GPS定位，为最近的运维人员推送抢修任务。工单派发后，系统进一步利用多模态数据优化维修过程。现场维修人员可使用智能终端获取用户语音记录、图片解析结果以及远程诊断数据，并在维修完成后上传设备修复照片或视频，系统利用OCR和图像识别技术对比维修前后的设备状态，确认问题是否彻底解决^[4]。一旦系统检测到仍存在异常，工单将自动进入二次复检流程，确保维修闭环管理。

2.4 多模态客服情感分析与交互优化

客服调控系统在客服交互过程中，采用语音情感分析、文本情绪识别以及面部表情分析技术优化客户体验。系统在语音交互时，利用声学模型分析用户语调、音量以及语速波动，判断是否存在焦虑或不满情绪。例如，若用户在通话过程中语速加快且音量上升，系统会自动标记

对话为“高情绪波动”，并动态调整回复策略，提供更详细的解释或直接转接至高级客服。而在文本交互中，系统运用自然语言处理模型分析用户输入的语句情感倾向，若检测到“投诉”“不满意”等负面关键词，则优先安排人工客服介入，避免固化的机器回复加剧用户不满。在视频交互模式下，客服终端利用面部表情识别技术分析用户情绪状态，若检测到用户在自助操作过程中表现出困惑或烦躁，系统会主动提供简化操作步骤或引导至视频客服模式的建议，以提升交互体验^[5]。综合多模态情感分析结果后，客服系统还能针对不同类型的用户构建个性化交互模型。例如，系统检测到用户在过去多次交互中表现出高情绪波动，未来的客服对话中将优先提供更直接的处理方案，减少不必要的重复解释。

3 实践案例分析

客服调控系统作为国家电网的核心客户服务平台，承担着海量客户咨询、投诉和业务办理的任务。随着客户需求的多样化和复杂化，传统的单一交互方式已难以满足高效、精准的服务要求。为此，系统引入了多模态交互技术，构建智能排班解决方案。该方案整合语音识别、自然语言处理与机器人流程自动化技术，实现从语音输入到自动排班的全流程智能化。

多模态交互技术在客服调控系统中落地实施，具体包括以下核心环节。（1）语音采集与识别。系统支持客服主管通过语音输入排班指令，采用物理按键触发语音采集，避免因环境噪声造

表 1 多模态交互技术具体应用成效

指标	具体效果
语音识别准确率	普通话场景下可达 98.7%，方言适配率达 89.5%
排班操作成功率	单次排班任务执行成功率 99.1%，批量排班（100 人）成功率 97.8%
自动化效率提升	相较人工模式，单次排班平均耗时从 38min 缩短至 2.7s，效率提升 99.88%
容错机制表现	自动重试策略成功挽回 68% 的临时性故障，人工干预率降低至 0.9%
系统稳定性	在 7×24 小时连续运行测试中，CPU 占用率最高不超过 78%，关键服务可用性达 99.999%

成误触发。输入语音经麦克风采集后，利用先进的端到端语音识别模型进行实时转写。该模型结合自适应降噪技术，优化不同背景噪声下的识别性能，并支持中英文混合识别以及电力行业术语解析。在高噪声环境下，系统识别准确率可达 92.3%。（2）语义解析与意图识别。转写后的文本由指令解析层处理，采用 Transformer 架构的自然语言处理模型解析语义，并结合上下文追踪技术，确保跨轮次指令的连贯性。系统能够精准提取排班相关要素，如日期、时间、班次类型、人员调整需求等。同时，为增强对模糊指令的理解能力，引入 BERT 意图识别模型，使系统能正确解析如“把今天晚班的张三调整到明天白班”这类复杂指令，并给出合理的排班方案。（3）自动排班执行。解析出的排

班指令由 RPA 执行层自动处理，系统通过调用排班管理平台 API 或模拟人工操作完成相应任务。RPA 采用动态脚本加载机制，支持灵活配置不同班次规则，如轮班、紧急替班等。同时，系统内置异常自愈机制，能在排班执行过程中识别异常情况，如系统界面未响应、数据录入错误等，并自动进行纠正或重试，确保排班任务的稳定执行。

在真实客服场景中，该系统进行了一系列性能评估与应用验证。多模态交互技术具体应用成效如表 1 所示。

虽然当前系统已显著提升了排班效率，但仍存在改进空间。例如，方言识别的适配性仍需优化，特别是在南方方言环境下，语音转写准确率仍低于业务可接受阈值。此外，RPA 执行层对目标系统界面变更较敏感，未来可探索更具适应性的自动化适配方案。针对长语音输入识别精度下降的问题，后续可通过优化流式处理机制提高系统鲁棒性。

结语

本文系统分析了客服调控系统中多模态交互技术的应用模式，具体涵盖语音、文本、图像以及视频的融合应用。分析显示，多模态技术在智能问答系统中通过语音识别与文本分析，提高了客服交互的精准度；在远程故障诊断中，借助计算机视觉及增强现实技术，实现了高效的远程辅助；在智能工单调度方面，整合多模态数据优化资源匹配，提高了抢修工作的精准性；在客服情感分析环节，利用语音与面部表情识别，实现了情绪感知与个性化响应。未来，客服调控系统可进一步优化多模态交互技术，提升复杂环境下的语音识别准确率，并强化情感分析模型，以实现更具人性化的智能客服体验。

引用

[1] 殷冀,黄凯.信息产品设计中的多模态交互与用户认知探析[J].数字技术与应用,2024,42(7):167-169.

[2] 鲁特刚.多模态人机交互技术研究[D].西安:西安电子科技大学,2023.

[3] 杨思琪.多模态交互技术在数字体验设计中的应用[J].集成电路应用,2024,41(12):32-33.

[4] 吴柯蓉.基于大数据分析方法的电力服务质量提升研究[D].南宁:广西大学,2021.

[5] 任立国,刘翠,张乐.快速响应客户诉求确保服务精准高效——国网客服中心多措并举服务抗洪保电[J].农村电工,2020,28(10):13.

