

语音识别在值班系统中的应用研究

周羽丰¹ 唐丽²
ZHOU Yufeng TANG Li

摘要

从提高值班系统智能化水平入手，在分析语音识别技术的基础上探讨语音识别技术在值班系统中的应用场景。借助IP语音调度系统的话音交换存储能力，基于值班系统设计一套话音自动解析转写的应用方案，制定一套值班话音数据交互流程及交互标准协议，实现语音识别技术在值班系统中的应用落地。

关键词

值班系统；语音识别；语音调度

doi: 10.3969/j.issn.1672-9528.2024.08.005

0 引言

随着信息化建设的不断完善，值班室也进行了信息化建设改造，指控系统、视频设备、语音电话等软件和设备在值班工作中扮演着越来越重要的角色，使得日常值班作业的整体效率得到明显提升。但随着信息化手段的不断丰富，值班人员的工作负荷也日益加重。值班室的电话会议、电话抽查、电话指示以及电话报告各类突发事件等，都需要值班员花大量精力用在人工记录电话纪要、登记值班事件上。因此，为值班员提供高效准确的话音转换功能是值班系统当前急需解决的问题之一。

随着人工智能技术的发展，智能语音识别技术从理论走向应用。近年来，随着语音识别技术的不断成熟，简单场景下的语音识别准确率能够达到97%以上。借助语音识别技术，突破传统值班事件手工登记模式的限制，使用语音识别转写代替人工登记，可以明显提高值班人员的作业效率，降低值班工作强度。同时，在深度学习的加持下，智能语音技术会针对处理内容和人工修正自动对语音识别结果进行优化，提升了语音识别技术在实际应用上的潜力^[1]。

1 需求背景

目前日常值班作业中存在大量的电话使用场景，例如上级通过电话下达指示命令、上级通过电话对下级进行值勤抽查、下级值班员通过电话上报突发事件情况等。在没有语音识别技术应用的情况下，通过人工记录电话信息的方式效率低且出错率高，若指示或值班事件情况出现记录偏差，容易导致延误事件处置时机的严重后果。

目前各级值班室里除了配备程控电话外，还配备了IP语

音调度单元，主要包括IP语音调度单元服务器、语音网关、网络录音存储设备、单元运维终端和IP语音调度终端。其以IP语音调度系统为核心，为值班室提供了一体化指挥调度功能。本文借助IP语音调度系统的话音交换存储能力，基于语音识别技术，制定了值班话音数据交互流程及交互标准协议，结合值班系统设计和开发了一套多用户话音解析转写记录软件，将语音识别技术真正运用到日常值班作业的场景中，实现对电话话音的实时转写记录、电话纪要自动生成管理，解决用户实际使用中的痛点。

2 语音识别技术

2.1 语音识别技术概述

语音识别技术是实现人机交互的关键一环，它的功能就是通过计算机将信息完成从“声音”到“文字”的处理转换，将语音信号转变为相应的文本或命令。语音识别是语音信号处理的重要研究方向，属于模式识别的重要分支，同时也涉及声学、生理学、信号处理、语言学、心理学等多学科知识的交叉融合，是一项集声学、语音学、计算机、信息处理、人工智能等于一身的综合技术，在计算机、通信领域等都有较为广泛的应用前景。

语音识别技术在“能听会说”的智能系统中扮演着重要角色，相当于给计算机系统安装上“耳朵”，使其具备“能听”的功能，进而借助计算机实现海量语音数据的快速处理。近年来，随着人工智能的兴起，语音识别技术开始从实验室走向市场，已逐渐走进我们的日常生活。现在语音识别已用于许多领域，提供如自动客服、自动语音翻译、命令控制、语音验证码等多项应用。

2.2 语音识别基本流程

从系统结构上来说，语音识别系统的本质是将输入的音

1. 中国电子科技集团公司第二十八研究所 江苏南京 210000

2. 32369 部队 北京 100042

频流信号借助训练好的声学模型和语言模型转变为文本信息等识别结果的过程,一般主要由预处理、特征提取、识别解码、模型库、识别结果处理等模块构成^[2],如图1所示。

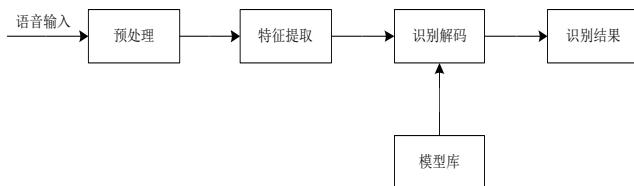


图1 语音识别基本流程图

步骤1：语音预处理。为了改善语音效果,保证系统接收的语音质量,需要对输入的语音信号进行预处理过程,预处理过程主要包括去除杂讯、高频信号加强等。

步骤2：特征提取。从经过预处理后的语音信号中提取具有辨识性的成分,使得语音模拟信号数字化,并存储为计算机可以处理的格式。

步骤3：识别解码。特征提取完毕后送入解码器,解码器根据事先训练好的声学模型及语言模型对测试特征进行模式匹配。

步骤4：识别结果输出。最后输出似然度最高的词序列作为识别结果。

3 典型应用设计

3.1 设计概述

目前,从事语音识别相关应用研发的厂商较多,像科大讯飞、百度等都提供了有较成熟的语音识别解决方案,但语音数据的交互参数协议不统一。IP语音调度系统主要有融智通、广有和佳讯飞鸿三个厂家,各家提供的话音数据交互接口形式也是多种多样。从值班系统的兼容性、稳定性、可维护性等角度考虑,需要针对值班系统与IP语音调度系统和语音识别系统的话音数据交互需求,设计一套标准的数据交互流程,并分别制定相应的交互协议规范,以适应控制接入不同厂家的语音识别服务和语音调度服务。

3.2 适用场景

(1) 值班抽查 / 指示下达 / 事件上报

上级值班员采取电话询问的方式,对下级执勤值班、警卫执勤和应急分队人员在位履职、执行任务等情况进行不定时检查,督导值班制度落实,检验快速反应能力。若发现问题,则及时通报整改。值班抽查过程全程记录生成抽查情况纪要,便于管理追溯。上级通过电话向下级值班员下达命令指示,下级值班员登记生成指示纪要。下级值班员通过电话将突发事件进展情况/处置情况上报上级值班员,上级值班员进行事件情况登记备案,如图2所示。

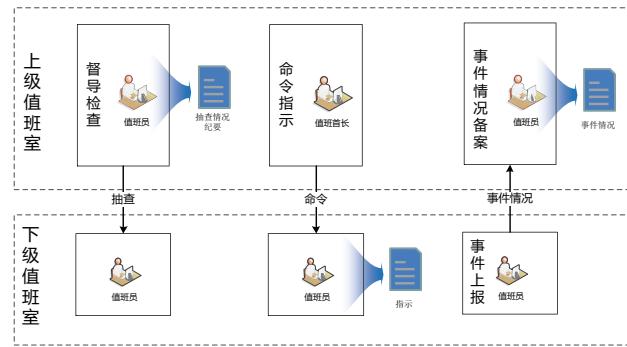


图2 值班抽查 / 指示下达 / 事件上报场景示意图

(2) 专项任务会议讨论

上级各值班要素针对各要素专项任务向下属发起电话会议讨论,上级安排分配任务,下级报告任务执行困难需求。在上级值班系统终端生成电话会议流水纪要,供上级值班员管理追溯,如图3所示。

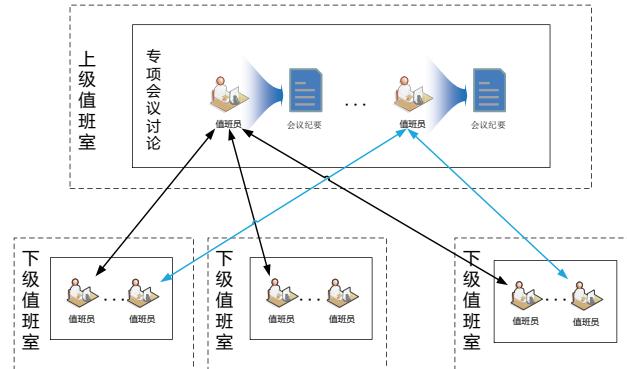


图3 专项任务会议讨论示意图

3.3 交互流程设计

3.3.1 用户信息接入与同步交互设计

在多用户通话场景下,通过声纹识别的方式解析多路混合话音数据流的准确率低。为了确保值班电话话音数据的高识别率,需要将获取到的每一路话音媒体流加上电话号码等相关用户身份信息,以便系统进行单路话音识别转写后的电话纪要合成。因此,首先要解决用户信息的随机接入及上下级用户信息同步的问题,要求值班系统能够控制IP语音调度服务进行指定话音终端的用户上线注册与下线注销,并能够自动同步上下级用户信息。具体用户信息接入及用户信息同步交互流程见图4所示。

(1) 系统开通后,值班系统自动抽取平台用户信息,值班员通过值班系统配置绑定用户对应的电话号码、角色要素、所属部门等信息后,值班系统将用户信息进行上下级自动同步。

(2) 值班员通过值班系统配置绑定本级值班室内各席位终端和话音终端。

(3) 用户在席位终端（客户端）登录值班系统，值班系统获取登录用户信息及登录客户端 IP。

(4) 值班系统通过配置的绑定信息, 获取上线用户对应使用的话音终端IP和电话号码, 并向IP语音调度服务推送。

(5) IP 语音调度服务根据推送的用户上线注册信息,向指定话音终端推送注册电话号码, 实现话音终端设置伴随用户移动接入。

(6) 当用户从值班系统中下线时, 值班系统获取下线用户信息及下线客户端 IP, 获取对应需要下线的话音终端 IP, 由 IP 语音调度服务向指定话音终端推送注销指令, 实现用户下线与话音终端号码注销。

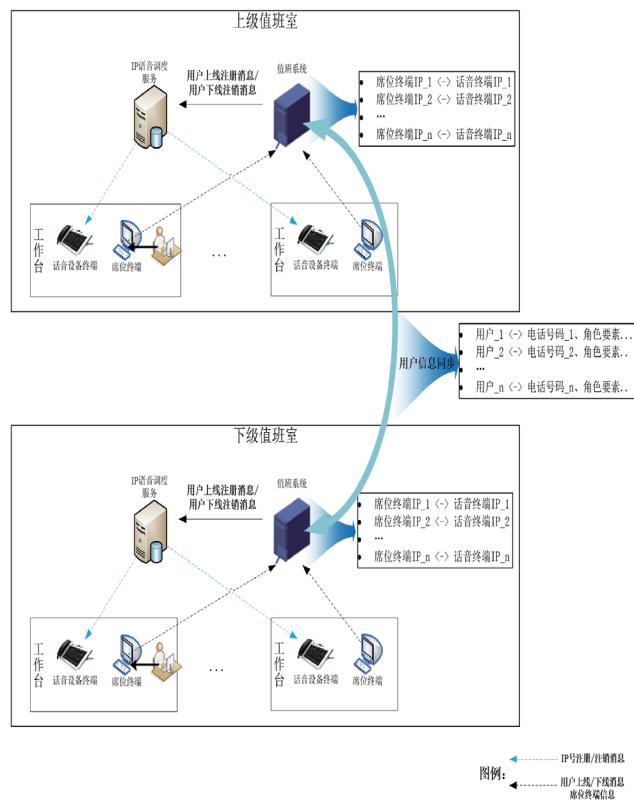


图 4 用户信息接入与同步信息交互示意图

3.3.2 话音数据调度识别交互设计

电话会议场景下，需要值班系统根据订阅用户信息，向IP语音调度服务订阅指定话路的话音数据，调用语音识别服务解析话音数据，并根据话音的来源用户及时间先后顺序，生成电话会议流水纪要。具体话音数据的调度识别交互流程如图5所示。

(1) 用户通过值班系统订阅话音数据, 值班系统通过在线用户信息获取对应绑定的电话号码, 向 IP 语音服务订阅与该电话号码通话的相关话路话音数据。

(2) IP 语音调度服务根据订阅信息, 向值班系统推送指定话路的话音数据、时间戳信息、电话号码信息等。

(3) 值班系统从推送信息中抽取语音数据，向语音识别系统发送。

别服务发起语音识别请求，语音识别服务将音频流数据实时转换成文字流数据结果，并返回给值班系统。

(4) 值班系统向订阅用户所在的指定席位终端(客户机)推送话音解析文字数据, 实现即时对语音内容的识别, 展现对应文字内容, 生成电话纪要。

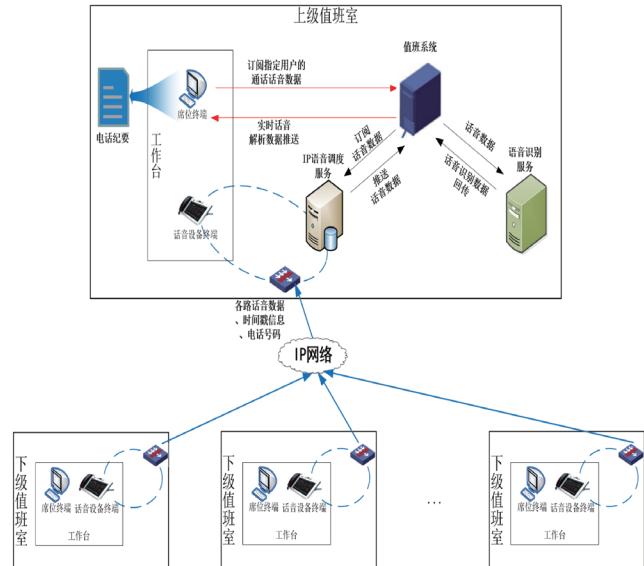


图 5 语音数据调度识别交互示意图

3.4 交互协议设计

针对 IP 语音调度服务，对用户信息注册 / 注销控制协议、话音数据订阅协议和话音数据推送协议进行规范，各 IP 语音调度服务提供商根据协议进行接口开发，使得 IP 语音调度服务能够根据用户上 / 下线同步进行用户信息的注册 / 注销，能够根据订阅信息推送指定话路话音数据。针对语音识别服务，对识别请求协议和分析结果查询协议进行规范，各语音识别服务提供商根据协议进行接口开发，使语音识别服务具备建立连接、接受任务、执行任务、接受进度查询、识别结果回传的功能。

值班系统主要负责用户信息注册/注销控制、用户语音数据推送控制、语音识别任务发起、识别结果查询和最终结果的综合呈现。针对用户交互，设计统一的前端展示页面，用以对任务下发、执行进度、识别结果等内容进行统一的展示。

所有交互协议均采用 HTTP 协议，使用 POST 方法传输实体信息，考虑到不同运行环境的兼容性，协议内容均采用 UTF-8 字符集进行编码。请求地址规则为“`http://【服务地址】/【接口】?【公共参数】`”，其中服务地址指的是值班系统服务/语音识别服务/IP 语音调度服务部署的 IP 地址及端口，公共参数主要包括厂商标识和用户 ID 等通用参数。实体信息参数格式全部为 JSON 字符串对象。具体交互关系见表 1 所示。

表 1 交互关系说明

序号	A	B	信息交换方向	交互内容	协议标准	备注
1	值班系统	IP 语音调度服务	A→B	用户号码注册 / 注销信息	http 协议	—
2			A→B	话音订阅信息	http 协议	—
3			A←B	话音数据	http 协议	—
4		语音识别服务	A→B	话音识别任务	http 协议	—
5			A←B	话音识别结果	http 协议	—

3.4.1 与 IP 语音调度服务的交互协议

(1) 用户信息注册 / 注销控制协议

控制指定话音终端注册 / 注销用户对应的电话号码，实现与值班系统上 / 下线同步。

表 2 参数说明

序号	名称	类型	是否为必需值	说明
1	phoneID	String	是	话音终端 ID 号
2	phoneNum	String	是	用户绑定的电话号码
3	type	Int	是	类型 (0,1) , 0 表示用户上线注册, 1 表示用户下线注销
4	UserName	String	否	用户名
5	duty	String	否	职务
6	element	String	否	值班要素

(2) 话音数据订阅协议

向 IP 语音调度服务订阅与某指定电话号码通话的话路话音数据。

表 3 参数说明

序号	名称	类型	是否为必需值	说明
1	phoneID	String	是	话音终端 ID 号
2	phoneNum	String	是	电话号码

(3) 话音数据推送协议

IP 语音调度服务向值班系统推送不同话路 (电话号码) 的话音流数据。

表 4 参数说明

序号	名称	类型	是否为必需值	说明
1	phoneNum	String	是	电话号码
2	startTim	String	是	开始事件
3	endTim	String	是	结束时间
4	duration	Int	是	时长
5	voicedata	byte[]	是	话音流数据

3.4.2 与语音识别服务的交互协议

(1) 识别请求协议

值班系统向语音识别服务发起话音识别任务，设置全局

任务 ID，按照从 IP 语音调度服务获取到的话音数据顺序，依次向语音识别服务推送待解析话音流数据。

表 5 参数说明

序号	名称	类型	是否为必需值	说明
1	sessionId	String	是	会话任务 ID，全局系统会话必须唯一
2	xh	Int	是	话音数据顺序
3	sr	Int	是	采样率，每秒采样次数，默认 16 000
4	bps	Int	是	采样精度，单个采样点大小，默认 16
5	fs	Int	是	每包字节数，默认 4096
6	voicedata	byte[]	是	话音流数据

(2) 识别结果查询协议

语音识别服务根据协议接收到识别任务请求后，按序解析话音流数据，并依次向值班系统按协议推送话音识别结果数据。

表 6 参数说明

序号	名称	类型	是否为必需值	说明
1	sessionId	String	是	会话任务 ID，全局系统会话必须唯一
2	xh	Int	是	话音数据顺序
3	IdentRes	String	是	话音识别结果

4 总结

本文从提高值班系统智能化水平入手，借助语音识别技术和 IP 语音电话系统，设计实现了语音识别技术在值班系统中的应用落地，解决了用户实际使用中的痛点问题。后续随着语音识别技术的进一步发展成熟，通过日常通话的话音声纹数据，训练采集用户声纹信息并与用户身份绑定，可实现对程控录音电话的多路混合话音的高解析准确率，进一步提高语音识别技术在现代化指挥信息系统中的推广价值，丰富语音识别技术在指控系统中的应用场景。

参考文献：

- [1] 汪霜玲, 金欣, 王晓璇, 等. 指挥信息系统智能化发展能力演化路线 [J]. 指挥信息系统与技术, 2019, 10(3):46-49.
- [2] 王煊, 王冠, 蒋伟煜, 等. 用于管制语音理解的语义分析方法 [J]. 指挥信息系统与技术, 2019, 10(1):32-36.

【作者简介】

周羽丰 (1993—)，男，上海浦东人，本科，工程师，研究方向：信息技术。

唐丽 (1981—)，女，山西朔州人，硕士研究生，工程师，研究方向：信息通信。

(收稿日期：2024-05-22)