

本文引用格式: 郑旭琪.融合语音技术与声纹认证的电力通信设备管理系统[J].自动化与信息工程,2023,44(6):46-51.

ZHENG Xuqi. Power communication equipment management system integrating voice technology and voiceprint authentication[J]. Automation & Information Engineering, 2023,44(6):46-51.

融合语音技术与声纹认证的电力通信设备管理系统*

郑旭琪

(广东电网有限责任公司揭阳供电局, 广东 揭阳 522000)

摘要: 随着电网设备规模的不断扩大, 设备管理系统应用需求也不断地增长。针对传统人工管理效率低的问题, 提出融合语音技术与声纹认证的电力通信设备管理系统。首先, 构建面向电力通信设备词汇的语音语料库和声学词表; 然后, 采用深度神经网络——隐马尔可夫模型 (DNN-HMM) 识别语音; 最后, 结合声纹认证技术, 对操作人员的身份进行识别。该系统实现了从手动向语音智能交互操作的转变, 使设备管理系统更加高效安全。

关键词: 智能电网; 设备管理系统; 语音交互; 声纹认证; 语音语料库; 声学词表; DNN-HMM

中图分类号: TN912.3

文献标志码: A

文章编号: 1674-2605(2023)06-0008-06

DOI: 10.3969/j.issn.1674-2605.2023.06.008

Power Communication Equipment Management System Integrating Voice Technology and Voiceprint Authentication

ZHENG Xuqi

(Jieyang Power Supply Bureau of Guangdong Power Grid Corporation, Jieyang 522000, China)

Abstract: With the continuous expansion of power grid equipment, the demand for equipment management system applications is also constantly increasing. Aiming at the problem of low efficiency in traditional manual management, a power communication equipment management system integrating voice technology and voiceprint authentication is proposed. Firstly, build a voice corpus and acoustic vocabulary for power communication equipment vocabulary; Then, deep neural network-hidden Markov model (DNN-HMM) is used to recognize voice; Finally, combined with voiceprint authentication technology, the operator's identity is identified. The system has achieved a transition from manual to voice intelligent interactive operations, making the device management system more efficient and secure.

Keywords: smart grid; equipment management system; voice interaction; voiceprint certification; voice corpus; acoustic vocabulary; DNN-HMM

0 引言

随着中国经济的快速发展, 电力资源的需求也日益增长, 建设统一而高效的智能电网成为中国电网发展的重要方向^[1-3]。电力通信设备管理系统是智能电网的核心, 现有的管理方式依赖于人工记录和手动操作。在面对庞大的设备数量和复杂的操作流程时, 人工管理方式存在效率低和耗时长等问题。随着电网安全化和智能化的需求越来越高, 语音人机交互成为关键技术之一^[4]。

目前, 深度学习在语音识别领域已取得显著成果, 使语音识别系统的性能不断提升, 为人机交互、自然语言处理等技术带来许多创新应用^[5-7]。文献[8]提出一种基于深度神经网络 (deep neural networks, DNN) 的电力系统调控语音识别方法, 并能对四川方言进行训练优化。文献[9]提出一种基于 DNN 的电力调度语音识别方法, 针对电力调度领域建立语音语料库, 应用 DNN 和隐马尔可夫模型 (hidden Markov model, HMM) 进行声学模型训练, 实现端点检测、语音转文

46 * 基金项目: 广东电网有限责任公司揭阳供电局职创项目
“一种离线式通信资料智能管理软件工具的研发”资助。

字等功能。文献[10]探讨了基于长短时记忆（long short-term memory, LSTM）神经网络的声纹识别方法，将语谱图作为声纹特征表达形式输入到 LSTM 模型进行训练，实现声纹个性特征的识别。目前相关研究主要集中在电力调度领域的语音识别方法，而将这些方法应用于电力通信设备管理系统的研究较少。考虑到电力通信设备管理系统信息的安全性，需要构建电力通信设备词汇的语料库，建立与系统相应的语言模型，并融合声纹认证技术提高系统的安全性。

因此，本文提出一种融合语音技术与声纹认证的电力通信设备管理系统。通过构建 DNN-HMM，实现电力通信设备词汇以及设备操作指令的准确识别；结合声纹认证技术，采用梅尔频率倒谱系数（Mel-frequency cepstral coefficients, MFCC）特征提取和高斯混合模型（Gaussian mixture model, GMM），实现用户身份的验证，避免非授权人员误操作，提高系统的效率和安全性。

1 系统框架

融合语音技术与声纹认证的电力通信设备管理系统主要包括语音识别子系统、声纹识别子系统和系统交互过程，系统框图如图 1 所示。

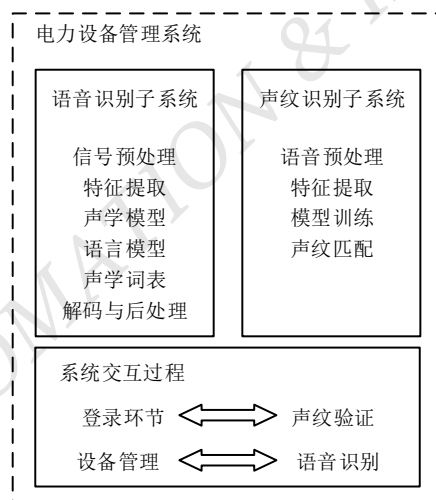


图 1 融合语音技术与声纹认证的电力通信设备管理系统框图

语音识别子系统负责识别语音指令；声纹识别子

系统用于识别操作人员的身份；在系统交互过程中，管理员通过声纹验证登录系统，利用语音识别技术对设备进行管理，两个子系统相互协作，提高了电力通信设备管理系统的智能化和安全性。

1.1 语音识别子系统

语音识别子系统主要包括信号预处理、特征提取、模型网络（包括声学模型、语言模型和声学词表）、解码与后处理 4 个部分。

信号预处理：去除语音信号中的噪音和干扰，并将语音信号分割成小帧，以便后续处理。

特征提取：利用 MFCC 从语音信号中提取数值特征，用于捕捉频谱特征，区分不同的音素。

模型网络：声学模型将语音特征与语音单元关联；语言模型建模语言结构；声学词表提供唯一标识符。

解码与后处理：结合声学模型和语言模型的输出，选择最可能的识别结果，进行后处理以提高系统识别的准确性。

语音识别原理如图 2 所示。

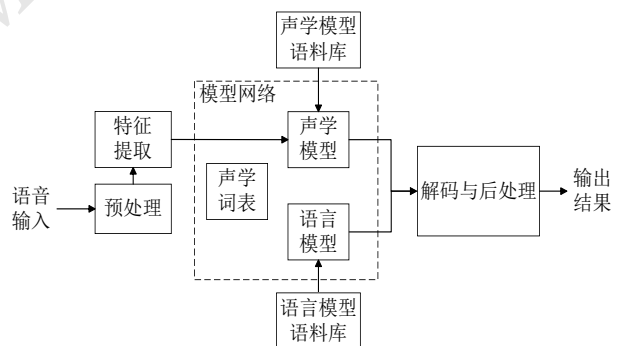


图 2 语音识别原理图

1.2 声纹识别子系统

声纹识别子系统主要包括语音预处理、特征提取、模型训练和声纹匹配，其基本原理是通过提取语音信号的声学特征，建立特征数据库，再通过匹配算法对语音进行识别。

语音预处理：对语音样本进行滤波、去噪、剪切语音片段等处理，提高语音信号的质量。

特征提取：利用线性预测系数（linear prediction coefficients, LPC）、倒谱系数（Cepstrum）、MFCC 等方法从预处理后的语音样本中提取能够代表个人声

音特点的声纹特征参数。

模型训练: 利用提取的声纹特征参数, 训练 GMM 声纹模型。假设一个人的语音由 M 个高斯分布混合生成, 其概率密度函数为

$$p(\mathbf{x}) = \sum_{i=1}^M \alpha_i g(\mathbf{x} | \mu_i, \Sigma_i) \quad (1)$$

式中: \mathbf{x} 为声纹的特征向量, α_i 是混合权重, $g(\mathbf{x} | \mu_i, \Sigma_i)$ 为第 i 个高斯分量。

声纹匹配: 对待测声纹进行似然值计算, 以确定其与已建模的声纹特征的匹配程度, 若高于阈值则匹配成功。匹配评分为模型对测试特征向量集 $\mathbf{X} = x_1, x_2, \dots, x_T$ 的似然值 $p(\mathbf{X} | \lambda)$, 即

$$S_{\text{score}} = \log p(\mathbf{X} | \lambda) = \sum_{i=1}^T \log p(x_i | \lambda) \quad (2)$$

通过以上步骤, 声纹识别子系统可以识别说话人的身份。

2 语音识别子系统模型网络构建

2.1 语料库建立

语料库对提升语音识别准确性和语义理解至关重要^[11]。语音识别系统中的声学模型和语言模型训练都基于语料库。语料库构成图如图 3 所示。

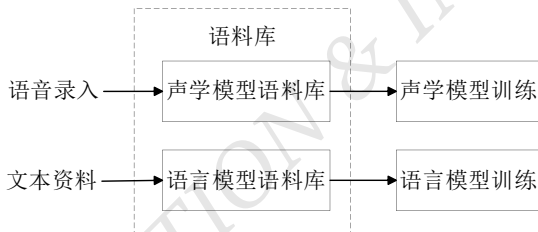


图 3 语料库构成图

电力通信设备管理系统的语料库通过 5 位 (2 男 3 女) 专业语音录入人员录入, 每位人员录入音频数据 1 000 段, 共录入 5 000 段。将这些音频数据划分为训练集、验证集和测试集。其中, 训练集包括 3 500 段音频数据, 用于训练声学模型; 验证集包括 750 段音频数据, 用于评估 DNN-HMM 的训练效果和调整参数; 测试集包括 750 段音频数据, 用于测试 DNN-HMM 的语音识别性能。音频数据参数设置如表 1

所示。

表 1 音频数据参数设置

参数类型	参数
采样率	16 kHz
比特率	16 bit
声道数	1
数据类型	Wav

2.2 声学词表建立

待测音频数据输入声学模型会将声音信号转为音素, 输入语言模型会将音素转为文本。声学词表是连接声学模型和语言模型的桥梁, 它包含了一系列音素与文本之间的映射, 使声学模型能够将声音信息转化为可被语言模型理解的文本形式。如当输入音频数据“你好”时, 其中文拼音是“ni3 hao3”, 经声学模型处理后其音素为“‘n’ ‘i3’ ‘h’ ‘ao3’”, 通过声学词表对应的音素转化为语言模型文本, 即识别出“你好”, 而不是同音词“拟好”。声学词表确保了声学模型和语言模型之间的有效沟通, 实现语音识别的准确性和可靠性。

为了满足电力通信设备管理系统的语言搜索功能, 需要创建一个专用的声学词表, 以确保系统能够准确理解用户的需求。首先, 准备与电力通信设备管理系统相关的查询操作语句, 并根据语法结构进行分词; 然后, 根据清华音素集数据库, 获得每个词的音素和拼音信息; 最后, 将音素与拼音相匹配, 构建完整的声学词表。声学词表文本格式如图 4 所示。

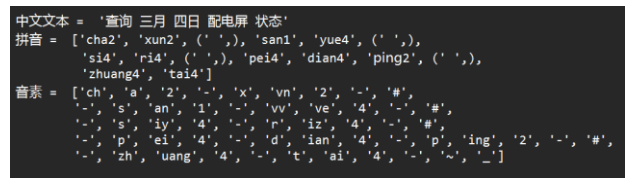


图 4 声学词表文本格式

2.3 DNN-HMM 构建

DNN 是典型的前馈神经网络, 通常包括输入层、隐藏层和输出层 3 部分^[12]。在 DNN-HMM 中, 输入层输入的声学特征 (如 MFCC 或声谱图) 表示为向量 \mathbf{x} ; 隐藏层包含多个神经元, 其输出可以表示为隐藏层

的激活向量 h ：

$$h = f(W_{hx} \cdot x + b_h) \quad (3)$$

式中： f 为激活函数， W_{hx} 为输入到隐藏层的权重矩阵， b_h 为隐藏层的偏置向量。

输出层与 HMM 的状态数相匹配，输出层的输出向量通常表示为 o ：

$$o = g(W_{oh} \cdot h + b_o) \quad (4)$$

式中： g 为输出层的激活函数； W_{oh} 为隐藏层到输出层的权重矩阵； b_o 为隐藏层的偏置向量，隐藏层的深度和神经元的数量可以根据任务的复杂性进行调整。

DNN 将声学特征 x 映射到每个 HMM 状态的概率分布表示为

$$P(state_i|x) = o_i \quad (5)$$

式中： $state_i$ 表示第 i 个 HMM 状态， o_i 表示输出层向量 o 中的第 i 个分量。

DNN-HMM 建模流程如图 5 所示。

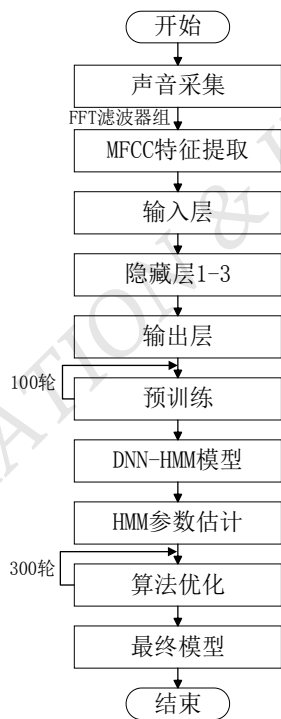


图 5 DNN-HMM 建模流程图

DNN-HMM 建模流程如下：

1) 采用基于 FFT 的滤波器组提取 39 维的 MFCC 特征，输入到 DNN 模型的特征维度为 $39 \times 11 = 429$ ；

2) DNN 模型包含 1 个输入层，3 个隐藏层和 1 个输出层，输入层节点数为 429 个，对应语音特征向量的维度；3 个隐藏层节点数均为 1 024 个；输出层节点数为 1 592 个，对应 HMM 的状态数；

3) DNN 采用 ReLU 激活函数，结合自适应矩估计 (adaptive moment estimation, Adam) 优化算法，提高模型的训练效率，训练过程利用 dropout 正则化，并采用“早停法”防止过拟合；

4) DNN 预训练迭代 100 轮后，将参数转移到 DNN-HMM，采用 Baum-Welch 算法估计 HMM 参数，再使用 EM 算法优化，再迭代 300 轮获得最终 DNN-HMM。

2.4 结果分析

利用词错率 (word error rate, WER) 来评估 DNN-HMM 的性能。DNN、HMM、DNN-HMM 3 种模型的词错率对比如表 2 所示。

表 2 3 种模型的词错率对比

声学模型	WER/%
DNN	22.31
HMM	32.05
DNN-HMM	19.23

在测试集上，DNN-HMM 的 WER 为 19.23%，优于 DNN 模型 (22.31%) 和 HMM (32.05%)，表明 DNN-HMM 能有效提取语音特征；利用 HMM 对语言时间序列进行建模，不仅提高了对电力通信设备名称的识别准确率，还能识别各种设备操作指令，实现了语音交互控制。

3 声纹识别子系统模型构建

3.1 声纹识别建模

声纹识别可以分为说话人辨认和说话人确认 2 种类型。说话人辨认是从一群人中确定某个测试语音属于某个人，说话人确认则是验证测试语音是否属于某一特定人^[13]。

实验中，从 50 位用户中随机选取 10 位作为管理

员进行注册。这 10 位管理员录制 20 个语音样本，用于 Enrollment 模型训练。声纹识别子系统模型构建流程图如图 6 所示。

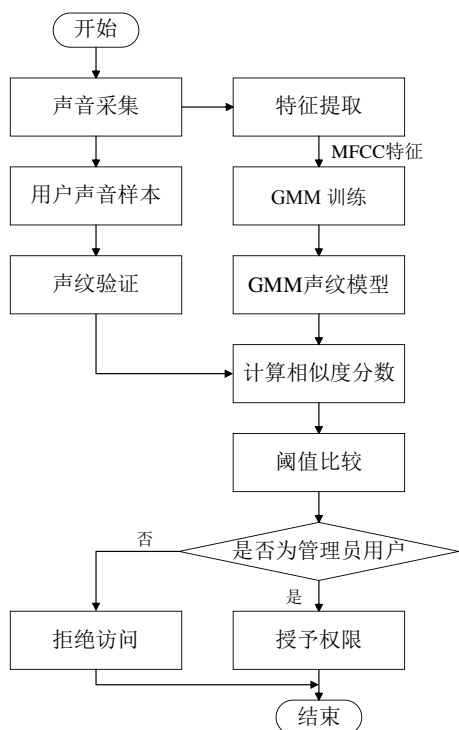


图 6 声纹识别子系统模型构建流程图

声纹识别子系统模型构建主要分为 4 个步骤：

- 1) 语音预处理：对录入的语音样本进行预处理，包括降噪/去直流等，提高后续处理效果；
- 2) 声学特征提取：通过对语音样本进行预加重、分帧、加窗、FFT 等处理提取 MFCC 特征；
- 3) GMM 训练：使用提取的 MFCC 特征，利用 EM 算法训练 GMM，提取声纹信息，获得 GMM 声纹模型，假设一个人的语音由 M 个高斯分布混合生成，其概率密度函数如公式(1)所示；
- 4) 声纹验证：输入测试语音的 MFCC 特征，计算与训练集中 Enrollment 模型的似然值评分，如公式(2)所示，若评分高于设定阈值，则表明测试语音与指定用户吻合，实现声纹验证。

通过声纹识别后，系统接受来自该用户的设备操作指令，从而保证系统操作的安全性。

3.2 测试结果及分析

在测试阶段，随机选取 5 位管理员，各采集 10 段

语音样本进行测试。对于 10 位测试用户的语音样本，系统能够正确识别出所有的管理员，假冒用户的误识别率为 8%。管理员的平均识别准确率达到 96%，验证了基于 MFCC 和 GMM 的声纹识别技术的有效性。管理员的语音样本验证评分明显高于阈值，而假冒用户的语音样本验证评分低于阈值，对于少量边界样本会出现误识别的情况。测试验证了基于 MFCC 特征和 GMM 的声纹识别技术，能够有效提取个人声纹信息，进行用户识别和验证，提高系统的安全性。

4 系统功能实现与测试

4.1 系统总体实现技术

电力通信设备管理系统界面如图 7 所示。

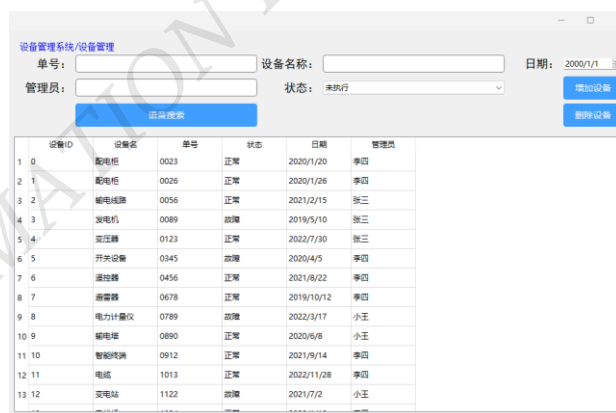


图 7 电力通信设备管理系统界面

在设备管理界面，用户可通过键盘输入设备信息，如设备名称、单号和管理员，查找设备的运行情况；还可通过语音搜索功能，直接使用麦克风录入语音指令，查询某设备的信息状态，系统通过语音识别技术将语音指令转化为文本，然后根据文本内容进行相关设备查询。

系统可以通过声纹识别进行登录，如图 8 所示。

用户使用声纹识别登录时，只需点击语音登录按钮，根据界面的提示，需要跟读特定的信息内容（本文采用一串数字信息）。系统通过声纹识别技术验证用户的声纹特征，并将其与预先注册的声纹特征进行比对，判断该语音是否与系统录入的身份一致。若验证成功，则用户可进入设备管理系统界面。



图8 声纹识别登录

4.2 语音搜索功能

在设备管理系统界面点击“语音搜索”，即可实现语音交互效果。用户说出具体指令，如“查询单号0089 整流屏”，麦克风采集语音信号后，通过语言模型识别成指令，根据文本查询语句，系统会匹配关键信息“单号”“0089”“整流屏”，通过设备数据库查找对应的设备，获取设备信息界面。语音搜索界面和搜索结果分别如图9和图10所示。



图9 语音搜索界面

设备ID	设备名	单号	状态	日期	管理员
1 5	整流屏	0089	正常	2023/2/15	张三

消息 查找到了!

OK

图10 搜索结果

作者简介:

郑旭琪, 男, 1987年生, 研究生, 工程师, 主要研究方向: 电力通信。E-mail: zqx20060909@163.com

5 结论

本文研究了语音识别技术及声纹识别技术在电力通信设备管理系统中的应用。在语音识别部分, 构建了针对电力通信行业词汇的语料库和声学模型, 实现了语音指令的识别转换。在声纹识别部分, 采用MFCC特征提取和GMM训练, 实现了对指定用户的声纹验证, 提高系统操作的安全性。

测试结果显示: 语言识别模型能够理解语音指令, 实现设备搜索功能, 提高设备管理的效率; 声纹验证能有效识别注册用户, 为系统访问提供安全保障。

参考文献

- [1] 张瑶,王傲寒,张宏.中国智能电网发展综述[J].电力系统保护与控制,2021,49(5):180187.
- [2] 潘磊,沈雪晴,黄文雯,等.基于云网协同的智能电网管理体系及应用分析[J].中国测试,2022,48(S1):211-217.
- [3] 戴诚,梁诗晨,肖静薇,等.多径成型因子在智能电网认知无线电信道干扰分析中的应用(英文)[J].机床与液压,2018,46(6):60-66.
- [4] 张四维,武永泉,秦涛,等.面向电力线路巡检的语音指令识别系统研究和应用[J].信息化研究,2021,47(5):6-12.
- [5] 刘文举,聂帅,梁山,等.基于深度学习语音分离技术的研究现状与进展[J].自动化学报,2016,42(6):819-833.
- [6] 顾亚文.人工智能深度学习背景下语音识别方法研究[J].软件,2022,43(5):122-124.
- [7] 王家,龙冬梅.深度学习在语音识别中的应用综述[J].电脑知识与技术,2020,16(34):191-192;197.
- [8] 胡翔,杨洋,蒋长江,等.一种基于深度神经网络的电力系统调度控制语音识别模型[J].电子器件,2023,46(1):90-95.
- [9] 窦建中,罗深增,金勇,等.基于深度神经网络的电力调度语音识别研究及应用[J].湖北电力,2019,43(3):16-22.
- [10] 刘晓璇,季怡,刘纯平.基于 LSTM 神经网络的声纹识别[J].计算机科学,2021,48(S2):270-274.
- [11] 单连飞,张越.电网调度专业语料库构建方法研究及应用[J].机械与电子,2022,40(4):73-76;80.
- [12] 贾嘉敏,程振,潘文林,等.基于 DNN-HMM 的伪语音声学建模[J].计算机时代,2022(8):61-64;68.
- [13] 张振国.面向个性化隐私保护的声纹生成方法研究[D].广州:广州大学,2023.