2022年英特尔杯大学生电子设计竞赛嵌入式系统专题邀请赛

2022 Intel Cup Undergraduate Electronic Design Contest

Embedded System Design Invitational Contest

**基于声纹识别的语声控制器**

**设计报告**



**项目成员**： 付东源、姚辰龙、朱子虚

**指导教师**： 孙文生

**参赛学校**： 北京邮电大学

**2022年英特尔杯大学生电子设计竞赛嵌入式系统专题邀请赛**

**参赛作品原创性声明**

项目团队郑重声明：所呈交的参赛作品报告，是项目成员（付东源、姚辰龙、朱子虚）独立进行研究工作所取得的成果。除文中已注明的引用内容外，本报告不包含任何其他个人或集体已经发表或撰写过的作品成果，不侵犯任何第三方知识产权或其他权利。项目团队完全意识到本声明的法律结果由项目团队承担。



**参赛队员签名**：

2022年7月24日

# **·摘要·**

**声纹**（Voiceprint）是语音中蕴含的、可表征和标识说话人的生物识别特征，以及基于这些特征和参数所建立的语音模型的总称。**声纹识别**（Voiceprint Recognition，VPR），又称说话人识别（Speaker Recognition），是一种根据语音信号中能够表示说话人信息的声纹特征，利用计算机以及各种信息识别技术，自动确认或分辨说话人身份的生物特征识别技术。

随着信息技术的飞速发展和各类智能终端的应用普及，生物特征识别技术广泛应用在各类身份验证场景。声纹作为一种长期稳定、隐私保护性强、不受距离限制、带有生理特性的特殊行为特征，其便捷性与安全性在众多认证方式中脱颖而出。本项目主要研究**声纹识别算法**和**语音识别技术**，并利用GNU-V40平台（下称“竞赛平台”）实现一个**具有身份认证功能的语声控制器**，该控制器可以接受特定人发出的指令，并驱动执行机构完成相应的操作。该项目要完成的具体任务如下：

为提升声纹模型训练过程的收敛速度和模型性能，提出一种Softmax交叉熵损失函数超参数优化算法，该算法通过在每次迭代时调整超参数比例因子和间距提升模型的收敛速度，找到网络达到性能最优的参数值，提升说话人网络模型的性能。为降低模型的资源利用率，使用深度可分离卷积替换一维扩张卷积，构建轻量级模型；使用ReLU6替换ReLU激活函数避免嵌入式平台运行时的精度损失；通过模型量化技术将模型中浮点实数以单字节整数格式存储，并充分利用硬件所支持的整数运算指令集，提升运算速度。

对于项目中的语音识别系统，基于主流开放平台的“短语音识别”服务进行快速构建。即通过调用网络接口，获取预先录制的说话人音频文件所对应的文本内容，进而进行语义分析。项目系统采集并存储说话人的语音信号样本后，声纹识别系统首先处理样本，通过与声纹数据库进行比对，确认说话人身份。系统则根据说话人身份确认结果，执行后续操作。

关键词：声纹识别，深度学习，身份认证，语音识别

# **·目录·**

（1）**项目背景**······························· 4

（1.1）**项目背景与意义**·························· 4

（1.1.1）项目背景··························· 4

（1.1.2）项目研究意义························· 5

（1.2）**声纹识别发展趋势**························· 6

（1.2.1）研究趋势：深度学习效果拔群，时变鲁棒成为焦点········· 6

（1.2.2）应用趋势：国标与商用并进，于实践中积累经验·········· 6

（1.3）**项目应用前景**··························· 6

（2）**项目设计方案**····························· 8

（2.1）**需求分析**····························· 8

（2.2）**系统架构与技术选型**························ 8

（2.3）**概要设计**····························· 9

（2.3.1）改进的时延神经网络文本无关声纹识别算法············ 9

（2.3.2）基于2.3.1中算法的声纹识别语声控制器系统 ··········· 12

（2.4）**详细设计**····························· 13

（2.4.1）综述····························· 00

（2.4.2）声纹识别用户终端······················· 00

（2.4.3）服务器···························· 00

（2.4.4）硬件网关··························· 00

（2.4.5）后台管理系统························· 00

（3）**项目系统测试**····························· 15

（3.1）**单元测试**····························· 00

（3.2）**集成测试**····························· 00

（4）**产品文档**······························· 00

（5）**项目总结**······························· 00

（5.1）**成员分工**····························· 00

（5.2）**开发日志**····························· 00

**附录** **参考文献与引用**··························· 00

**附录** **项目文件结构**···························· 00

# **1 项目背景**

## **1.1 项目背景与意义**

### 1.1.1 项目背景

（1）**传统生物识别方式漏洞频发，个人隐私泄露风险加剧：**

随着移动互联网设备的应用场景越发丰富，人们的生活与互联网变得密不可分。截至2018年12月，中国网民的规模达到了8.29亿，普及率攀升至59.6%。然而，计算机及网络技术也是一把“双刃剑”。万物互联的浪潮之下，网络身份认证问题日益凸显。

生物识别技术的突破，为人们提供了很好的解决方案。受益于近年来以大数据、云计算、深度学习为首的新一代人工智能技术的快速发展，生物识别技术，这一项以人体生理特征或行为特征作为身份标识来进行身份识别的技术在应用方面不断取得突破。以指纹识别及人脸识别技术为代表的传统生物识别方式，已在人们的生产生活中得到广泛应用。

然而，被“3·15”晚会曝光的人脸识别系统安全漏洞、熟人趁受害者不备使用其人脸与指纹完成验证、Deepfake诈骗兴起等在生物识别领域频发的安全事件，令生物特征数据原本作为优势而存在的“唯一性”这一特征，被发觉含有巨大风险。相较于数据在传输和认证过程中的安全漏洞，生物特征数据一旦被盗，大量带有唯一性的生物特征数据被盗取，给用户带来的风险将会更大。数据唯一，则意味着其不可撤销，也意味着一旦被泄露，被仿造，识别系统将存在崩溃的风险；而某些可能在用户不知情条件下被他人盗用的生物识别信息（如人脸、指纹等），也对个人隐私造成了严重的安全隐患。

（2）**生物识别样本形式从生理特征转向行为特征：**

生物特征识别的最大共性是唯一性——每个人都有独一无二的面部、指纹、虹膜等。由于每个人的生理特征具有与其他人不同的唯一性和在一定时期内不变的稳定性，故利用生物识别技术进行身份认定，相较于其他身份认证技术，具有更高的准确率。而将人体本身固有的、可唯一标记本人的生理和行为特征作为身份认证方式，也很好地解决了密钥的丢失和被破解等问题。

然而，生理特征的高度唯一性，使其被不法分子盗取、复现时对用户造成极为严重的安全风险。相比之下，基于行为特征的生物识别技术，因其样本隐私性弱、难以复现，且能体现用户意愿，具有更高的安全性。

人们认为，规避传统生物识别漏洞引发风险的一种可行方法，是基于生物行为特征的认证方法。自2017年起，智能语音交互技术发展加速迅猛，声纹识别技术逐渐成为最契合需求的选择。

声纹识别技术，作为一种基于带有生理特征的生物行为特征识别方式，相较于以指纹、人脸识别等技术为代表的完全基于生理特征的传统生物识别方式，具有隐私性弱、不怕丢失、难以伪造三大优势。

（1）**隐私性弱：**

声纹是一种具有生理特征的行为特征，其隐私性相对较弱，采集涉及到的用户隐私信息较少。跟读、语音提示等易于操作的样本采集方式，更易被用户接受。

（2）**不怕丢失：**

不同于指纹、虹膜、人脸等静态的生理特征，声纹作为一种动态的行为特征，不容易丢失，可以做到“失声（音）不失身（份）”。

（3）**难以伪造：**

对于不同的说话人，发音习惯和声纹特征是本人特有的，且在成年后基本不变。而同一人两次朗读相同的内容，也不可能发出完全相同的声音。声纹这种“大同小异”“蕴不变于千变万化之中”的特性，使其伪造难度极高。

（3）**后疫情时代 ，“非接触”识别带来新机遇：**

2020年2月，国务院在印发复工复产疫情防控措施指南的通知中提出“使用指纹考勤机的单位应暂时停用。”同月，由中国人民银行营业管理部制定的《北京市非银行支付机构复工复产防疫工作指引》提出要优化和丰富“非接触式服务”渠道和场景，强调疫情防控期间，暂缓人脸识别支付商户拓展。

针对疫情期间全民戴口罩时身份识别难度提高，以及公共场所进行身份认证时接触设备带来的交叉感染风险，如何在类似场景下准确识别用户身份、保护个人隐私和保障信息安全，是后疫情时代需要解决的重要问题。而目前广泛采用的人脸识别与指纹识别技术不能很好解决这一问题。摘下口罩进行人脸识别，将增大病毒空气传播、飞沫传播风险；按压指纹识别器进行指纹识别，可带来交叉感染风险。而声纹识别作为一种“非接触”识别方式，可直接阻断病毒传播途径，并间接降低疫情风险，在后疫情时代，具有重要应用价值。

为阻断病毒传播链条，机器人承接了消毒清洁、送药送餐、诊疗辅助等“一线工作”，VR看房、在线娱乐、在线教育等服务成为大众充实居家生活的必备之选，而打造远程银行、无人工厂的需求也比以往任何时候都更为迫切。这些“非接触式”服务的变革，不仅催生了新的经济模式——“非接触经济”，如在线办公、在线医疗等，还为声纹识别带来新的产业机遇。

（4）**语音交互个性化、声纹识别场景化**

目前，主流品牌智能手机系统应用与越来越多的第三方移动应用软件已配备语音交互功能，许多智能硬件厂商也在发力，积极布局以智能手机为中心，以语音交互为导向的智慧家居、智慧出行产品生态。语音识别与物联网技术赋能智能终端，将以往繁杂的操作变为一句句简单的语音指令，极大程度上方便了人们的生活。

声纹识别及其相关技术的发展，使语音交互的个性化、场景化解决方案不断涌现。如利用声纹确认技术完成个人日常生活中的各种事物访问控制的授权——声控锁屏、声控安全门、汽车声控锁等；利用声纹辨认技术，可支持智能音箱、智能语音助手等提供个性化服务，如针对家庭用户中的老年人、儿童等不同年龄段用户，按照兴趣推荐不同的歌曲、新闻，以及开放特定的功能权限等；利用声纹检出和追踪技术，可取代人工完成会议纪要，通过语音识别和声纹识别技术的结合，将会议录音通过语音识别技术识别说话内容、通过声纹识别技术标注每段话所对应的说话人，即可轻松完成多人会议记录，大大提高工作效率。

### 1.1.2 项目研究意义

项目实现改进时延神经网络文本无关声纹识别算法，并以该算法为核心，实现基于竞赛平台，软硬件结合的声纹识别系统，进而完成基于声纹识别的语声控制器的系统设计。综上所述，研究意义归纳如下：

（1）实现综述中的改进的时延神经网络文本无关声纹识别算法。考虑在有限资源硬件端部署的问题，对模型进行系统的优化，减少参数和计算量，使其更适合部署在更多低性能嵌入式平台。

（2）通过语音识别、语义识别，实现由说话人语音到控制外设指令的转换。

（3）打通多场景适用的，基于声纹识别的语声控制器系统的研发流程，将项目软硬件设计实现方案开源，作为声纹识别嵌入式开发的最佳实践案例，有效助力声纹识别及相关技术在企业、工厂、实验室等场景落地应用。

## **1.2 声纹识别发展趋势**

### 1.2.1 研究趋势：深度学习效果拔群，时变鲁棒成为焦点

随着深度学习的快速发展和应用，尝试将声纹识别与之结合的思路得到了广泛关注，各项相关研究效果拔群。Georg Heigold 等人提出了一种端到端的声纹确认方法，其取网络最后一层隐藏层的激活作为说话人表征，使用余弦距离判断两个表征向量是否为同一个说话人。Mirco Ravanelli 等人提出 SincNet 架构，以函数限定网络第一层卷积结构，让网络学习滤波器的截止频率，实现从原始语音信号直接学习，完成声纹识别任务。 Johan Rohdin 等人则模仿当前主流模型 i-vector-PLDA 模型的工作流，使用深度神经网络 DNN 实现工作流的每个部件，得到了不错的效果。

随着声纹识别技术逐渐成熟、趋于实用，与声纹识别相关的鲁棒性、安全性问题，也受到了研究和开发人员的关注，包括噪声、跨信道、多说话人、身体条件变化、说话方式变化、短语音等鲁棒性问题。

2000年至2010年，清华大学语音和语言中心对由于声纹随说话人年龄变化而发生变化从而导致系统识别性能下降的声纹时变问题进行了研究，提出了时变鲁棒的声纹特征；对使用录音和录音拼接攻击声纹识别系统这一安全问题进行了研究，并提出了切实可行的录音检测方法。

### 1.2.2 应用趋势：国标与商用并进，于实践中积累经验

为规范和正确引导声纹识别发展，国内已公布多项关于声纹识别的标准。2008 年，原信息产业部正式颁布实施了《自动声纹识别（说话人识别）技术规范》，这是我国第一个关于声纹识别的行业标准。2010年12月2日，公安部颁布实施了《安防声纹确认应用算法技术要求和测试方法》。2018年10月9日，中国人民银行正式对外发布《移动金融基于声纹识别的安全应用技术规范》金融行业标准。这是第一个被金融监管部门认可的生物识别标准，为声纹识别技术进入移动金融领域解决了标准难题。

声纹识别技术因其广阔的应用场景和巨大的潜在价值，从特定领域到民用领域，在国内外正迎来商用化浪潮。声纹识别首先在针对特定人群的国防安全、公安刑侦等领域投入使用，有力保障了国家和公共安全；国内外主流支付平台加持声纹确认技术，通过动态声纹密码的方式进行客户端身份认证，有效提升个人资金与交易支付安全性；互联网公司、智能终端厂商构建以“声纹识别+语音识别+物联网”为核心技术的智能家居、智慧生活、智能出行体系，令百姓生活更加便利。

## **1.3 项目应用前景**

根据《2019中国声纹识别产业发展白皮书》所述，声纹驱动控制器的未来发展方向，将不仅仅局限于安全保障。声纹特征的丰富性、稳固性与动态性，将使得其**在公共安全、移动金融、社会保险及个性化语音助手等技术领域的应用场景愈发多样。**

由于相关设备造价低廉，性能较好，声纹识别**在日常生活中应用前景广泛。**相关设备可用于智能家居，便捷百姓生活；可用于小区门禁系统，对进出小区住户进行记录；可用于银行自助取款机，快速识别取款人身份，为用户带来安全、便捷的使用体验。

**国防安全及公共安全等领域**是目前声纹辨认技术应用较为广泛的场景，如公检法人员能够通过通话记录中的声纹信息初步锁定嫌疑人，降低刑侦难度。公共安检处也可以通过在安检口设置声纹检测装置，初步筛查危险人物，提高公共保障安全性。

**移动金融方面**，以中国建设银行为首的一众银行逐步将声纹信息与动态密码相结合，通过多重身份认证的方式来提高安全性能。目前的主要实现方式是让用户读一段动态密码，将结果与用户事先认证好的声纹信息进行匹配，验证通过即可进行后续操作。

如今，**就准确率而言，声纹识别技术的识别准确率在理论上已超过90%**，但在现实生活中，说话人自身的独有特性、身体状况、年龄增长、情感波动等其他干扰因素，导致实际与实验中的理想值还存在一定偏差，仍可能出现对说话人身份产生误判的情况，因此可以提高准确率的方法还须进一步被发现、探讨、研究。但可以预见的是，在生物识别技术日趋成熟，生物识别设备加速普及的背景下，**声纹识别作为多生物特征识别融合技术体系中的关键一环，将在与其所应用领域业务形成发展的良性闭环同时，为信息时代的数字安全提供更为坚实的保障。**

# **2 项目设计方案**

## **2.1 需求分析**

（1）**快速声纹辨认能力：**

系统须识别说话人的靠近，给出说话人语音提示，并在说话人位于恰当距离处开始讲话时采集数秒的音频样本，通过将该样本与本地声纹模型数据库进行比对，确认某个声纹模型已存在于本地声纹模型数据库中用户的身份（已注册用户），或判断该用户的声纹模型不存在（未注册用户），并给出秒级响应。

（2）**快速声纹注册能力：**

需要时，系统可为用户提供注册功能。在提示下，令未注册用户与系统进行含有讲话、录音环节的交互，系统可处理用户讲话录音，生成并保存该用户声纹模型。此后，该用户可使用声纹识别完成身份验证。

（3）**根据声纹辨认与语音识别结果进行语义识别，执行简单指令：**

当系统辨认出用户为已注册用户时，须对其录音进行语音识别，提取文本信息。通过语义识别，将提取的文本信息映射至预设的简单指令集，系统据此决定并执行后续操作。

（4）**声纹管理能力：**

系统可为管理员提供交互界面，方便管理员进行系统配置、用户授权、声纹识别记录监控、声纹模型数据库管理等操作。

**用户**使用项目系统，主要完成**声纹注册识别、语音识别功能**；**管理员**使用项目系统，主要完成**声纹识别设备、声纹识别记录等管理功能。**鉴于用户与管理员的应用场景耦合度较低，且在实际应用时管理员可能需要管理若干声纹识别设备，故**在模块设计上也应将用户与管理员相关模块分离。**由此，采用基于B-S、C-S架构与终端的系统设计，分别实现系统的管理与注册/辨认功能。

## **2.2 系统架构与技术选型**

项目系统采用基于B-S、C-S架构与终端的系统设计，包含4个功能模块：**用户终端应用、服务器、****硬件网关、****管理系统。**

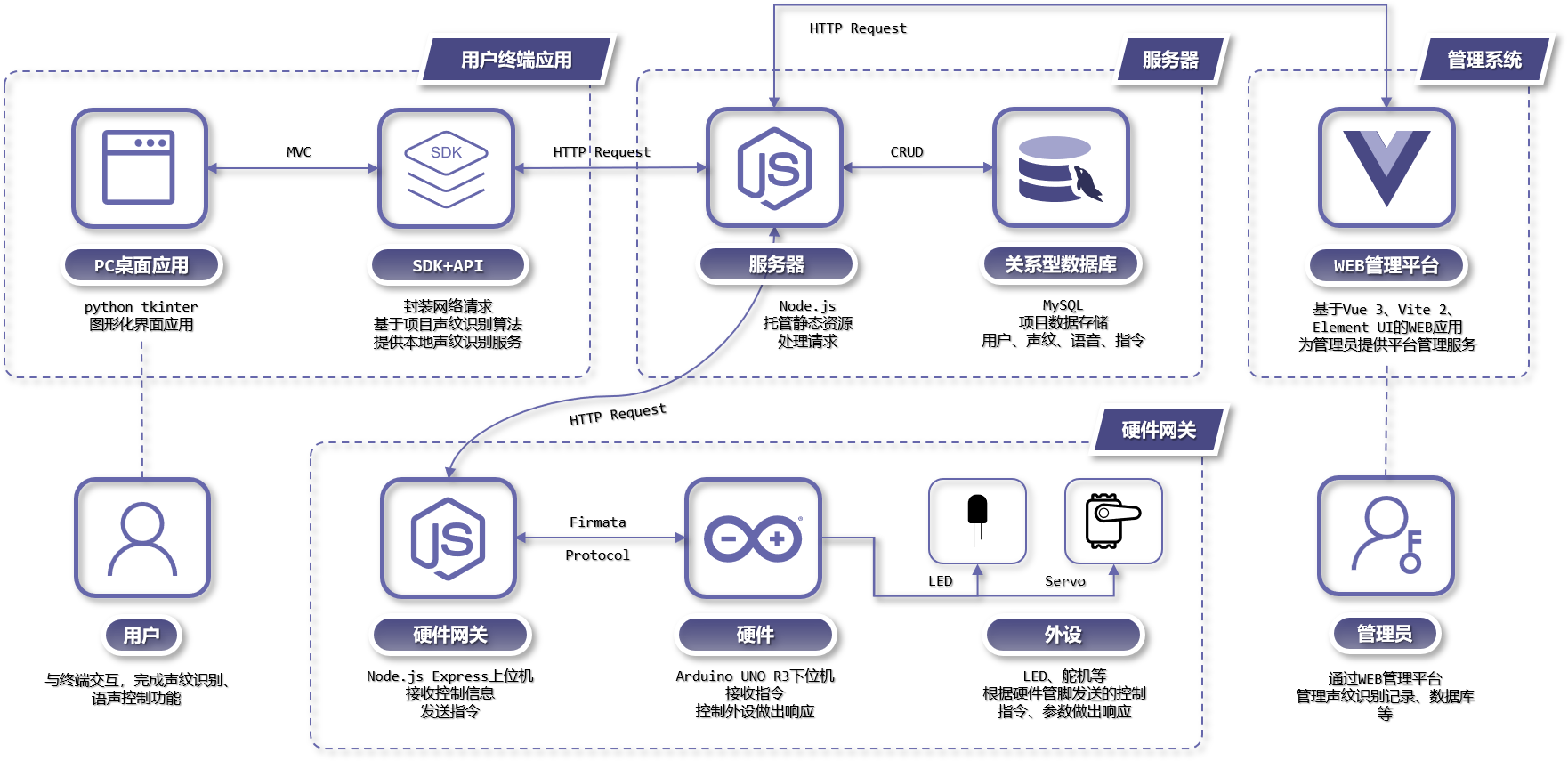
****

图 1 项目系统架构示意图

**（1）用户终端应用：**

基于python tkinter实现，面向用户，提供带有录音功能的，具有图形化界面的PC桌面应用。该应用通过集成包含项目声纹识别算法的软件开发工具包（Sonicredible SDK）并封装网络请求，为用户提供注册、声纹注册、声纹辨认与语声控制功能。

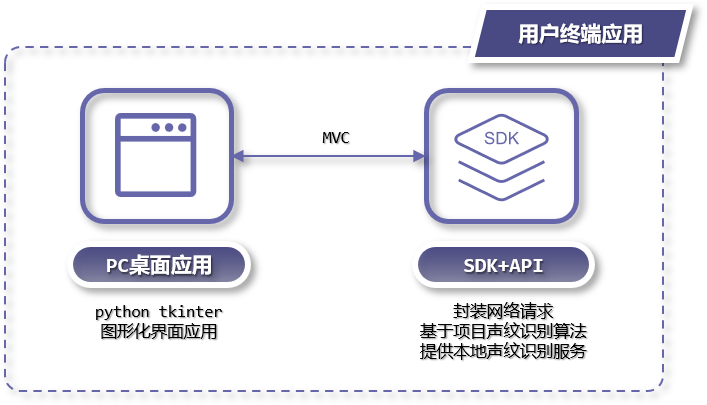


图 2 用户终端应用系统架构示意图

**（2）服务器：**

基于Node.js Express实现，使用MySQL关系型数据库，集成鉴权、注册、静态资源上传、托管等多种功能。作为项目系统功能模块的核心，为声纹识别用户终端应用、硬件网关提供HTTP RESTful API，进而实现数据传输与指令控制功能。

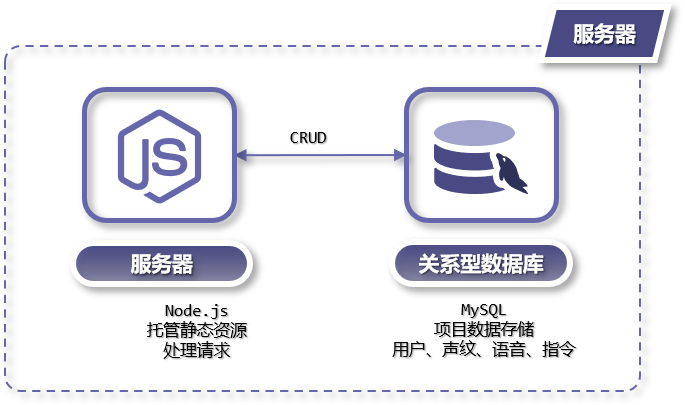


图 3 服务器系统架构示意图

**（3）硬件网关：**

基于Node.js Express构建上位机，以接有外设的Arduino UNO R3开发板为下位机，上位机作为网关，接收来自服务器的控制请求，通过Firmata协议，将控制指令发送至下位机，实现外设的控制。

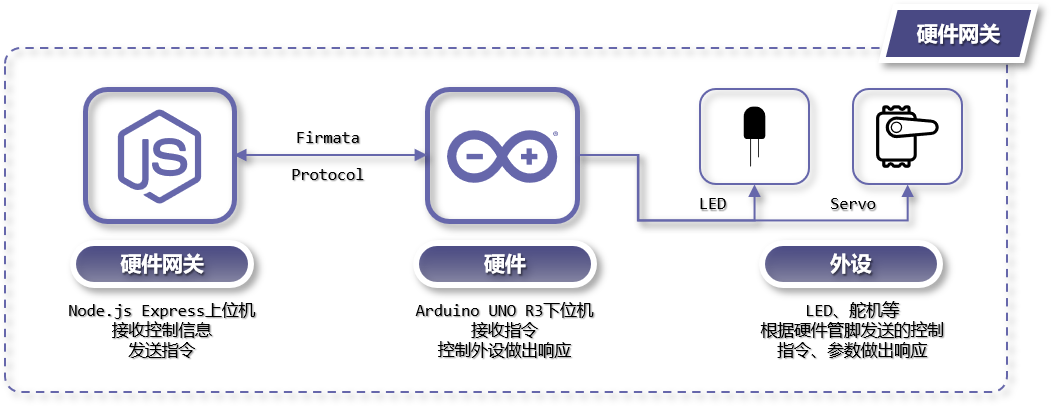


图 4 硬件网关系统架构示意图

**（4）管理系统：**

基于Vue实现，面向管理员，提供WEB应用。该应用通过集成用户管理、声纹识别音频管理、语义识别音频管理、音频指令管理等模块，为管理员监视系统状态、操作数据提供便利。



图 5 管理系统系统架构示意图

## **2.3 概要设计**

### 2.3.1 改进的时延神经网络文本无关声纹识别算法

**（1）算法功能：**

该算法尝试实现判定待测语音属于系统中已注册的说话人模型中的身份，即从数据库中找出是否有声纹与之匹配，是的身份辨认问题。完成任务的方法即设定一个声纹阈值，当待测语音与系统中所有人的相似度都低于该阈值，则认为待测语音对应的身份未在系统中注册过。该问题属于一项开集识别问题。

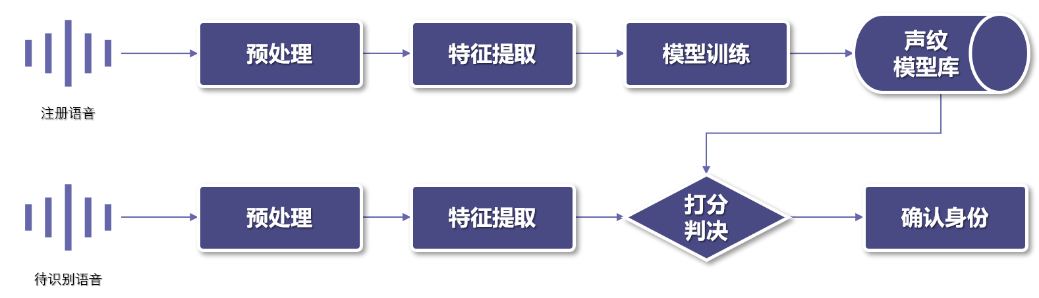


图 6 改进的时延神经网络文本无关声纹识别算法工作流程

算法首先需要实现声纹信息的数据库录入，即将注册者的语音信息进行预处理、特征提取；使用提取出的特征进行模型训练得出注册者的声纹并存入数据库。模型工作时，将收到的待识别语音预处理并提取出特征后，将特征与声纹模型库中的已有特征进行比对，使用打分算法判决确定说话人身份。

（2）**算法特征：**

根据系统功能框图，算法的部署分为三个模块：数据预处理，数据特征提取，模型训练及预测打分。

（1）**数据预处理：**

首先对训练集的数据进行了数据扩增。扩增方式参考如今多种较为流行的语音数据扩增方式，诸如加噪扩增（MUSAN，RIR）、变说话方式、On-the-fly 扩增、Voice Conversion（VC）和TTS扩增技术，最终采取加噪音混响、加背景音乐、倍速、频谱遮盖等方式作为数据增强的方式。对于每条语音信息，都从中随机选择若干类型进行扩增后投入模型训练。

对于待识别语音采集后的数据，为了最大程度避免失真，提高预测准确率，我们在预处理阶段分别采用**预加重**（Pre-emphasis）、**分帧**（Framing）、**加窗**等过程对其进行了降噪。

**预加重**过程采用一阶FIR高通数字滤波器补偿语音信号高频分量的损失，使信号的频谱变得平坦的同时，能够用同样的信噪比低频到高频求信号频谱，以便于更好地进行频谱分析或者声道参数分析。假设输入信号第时刻语音采样值表示为，经过预加重处理后的输出为。

**分帧**过程主要涉及两个参数：帧长（帧本身的长度）和帧移（帧与帧之间的间隔）。其基本思想是基于对语音信号的“短时平稳假设”以及希望更大限度的降低由于傅里叶变换带来的音频时域信息丢失，因此加入分帧处理过程，对于声纹识别系统将带来极大的提升。

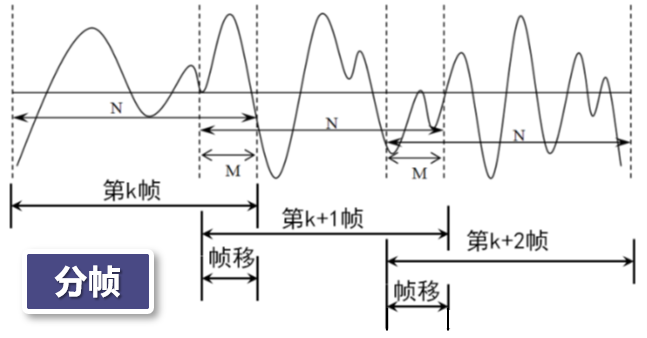


图 7 数据预处理——音频分帧

**加窗**过程的引入是为了解决分帧之后可能引起的吉布斯现象（Gibbs Phenomenon）与频谱泄露（Spectral Leakage）问题。即由于分帧产生信号的不连续性与截断效应，进而出现傅里叶变换后的频谱局部峰值，并且频域通带边界处出现多个旁瓣的现象。通过加窗处理，分帧后的每帧音频信号的每个值会被赋予不同权重。本项目中采用汉明窗（Hamming Window），此类型的窗函数中心到边缘权重值依次减小，可以有效减小分帧造成的信号不连续性。

（2）**数据特征提取：**

语音信号的特征提取方面，训练集与待识别数据的处理方式基本一致，均采用了基于梅尔频率尺度的语音信号特征提取技术，主要包括傅里叶变换、梅尔滤波器组及离散余弦变换。其目的在于对语音数据进行正弦提升（Sinusoidal Liftering），改善在噪声环境中的声纹识别效果。

（3）**模型训练与预测打分：**

项目所用模型，是参考ECAPA-TDNN模型后提出的一种基于改进时延神经网络的文本无关说话人辨认网络模型。预处理并通过声学特征提取得到的MFCC特征将输入构建好的神经网络模型，随后经过TDNN层和SE-ResNet2Block层扩展时间上下文提取帧级特征，加入SE-Block以明确对通道相关性进行建模，系统连接所有SE-ResNet2Block的输出进行多层特征聚合。此后，注意力统计池化层关注重要的帧级特征，计算加权平均值和标准差。最后，前馈神经网络提取音频段级特征，得到嵌入向量。

在对未注册声纹信息说话人的语音进行识别验证时，对特征向量进行采用余弦相似度（Cosine Similarity，CS）计算，选择相等错误率（EER）和最小检测代价函数（minDCF）作为评价指标。

（3）**算法改进：**

对损失函数与模型部署进行了优化。

（1）**损失函数优化：**

对于Softmax损失函数及其改进方案进行综合分析，发现它们都是通过在角度函数中引入各种类型的间距以增加类间距离。训练过程的收敛速度和模型性能在很大程度上取决于损失函数变体的超参数选择，这通常需要通过使用不同的超参数值多次重复训练来进行调整得到一个最优值。

为解决上述问题，提出对AAM-Softmax进行改进，使用**超参数随迭代次数动态调整策略**，该方法可以在迭代时动态调整超参数比例因子和间距，从而显著提高网络训练的收敛速度和说话人识别的精度。

分析损失函数，得出需要调制的超参数主要有两个，分别为和：参数对于狭窄的一些距离的范围进行了缩放，使得全连接层的输出结果更加可分；而参数增大了不同类别之间的间距，提升了分类的能力。这些超参数最终会影响到预测分类概率，理想的超参数的设置应该使得全连接层的输出覆盖的范围，并且在角度参数附近的梯度要大，使得训练更加有效。

对于尺度参数的动态调整方面，引入表示第次迭代时样本和其所属说话人类别的权重向量间的角度的参数，并使用迭代次数为时中全部个样本与目标样本类别权重夹角的中位数表示网络在当前批次上的优化程度。通过监控的相对大小，明确网络参数的优化状态，调整监督严格程度，经过多次迭代得到动态尺度参数。

类似地，对于参数的调整，定义全部个样本与所属说话人类别权重夹角的中位数，并使用此值对于优化程度进行表征，根据网络收敛情况调整间距。

（2）**模型部署优化：**

考虑到项目需要嵌入式部署与运行声纹模型，需要减少模型参数量与浮点运算次数，项目通过将**深度可分离一维卷积**（One-dimensional Depth-wise Separable Convolution，1DDSC）方案代替作为参考的ECAPA-TDNN模型使用的一维扩张卷积（Res2 DilatedCovlD）方案，实现系统性能的大幅改善。

同时，鉴于嵌入式设备通常采用float16或int8等较低精度的模型，ECAPA-TDNN 模型中作为激活函数使用的修正线性单元ReLU将存在精度损失。因此选择采用ReLU的改进版本——ReLU6，作为网络模型的激活函数，这样既保证了模型在嵌入式系统中正常的数值分辨率，也防止了“梯度爆炸”等情况的出现。

为了最大限度提升模型部署的资源瓶颈，项目引入模型量化技术。通过线性量化，将float 32权重矩阵用8位无符号整数（uint8）表示，大幅节省内存占用。项目团队还将结合大赛提供平台的嵌入式特性，对算法与模型进行进一步优化。

### 2.3.2 基于2.3.1中算法的声纹识别语声控制器系统

面向需求分析中的声纹注册、辨认、处理与管理等能力，将题述系统划分为3个子系统，分别为认证系统、命令系统、管理系统：

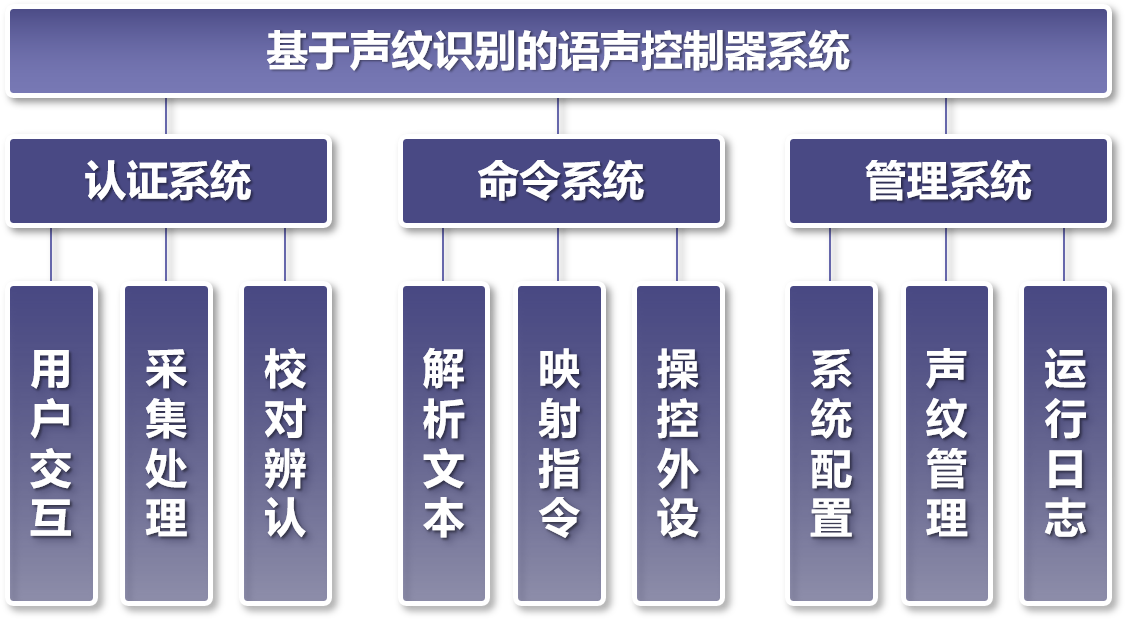


图 8 基于声纹识别语声控制器系统模块概要设计

（1）**认证系统：**

具有采集、处理、校对、辨认用户声纹信息及与用户交互的功能。

终端方面，接入外设，指引用户完成注册或验证操作，实现对用户语音信息的采集；接入网络，获取在线声纹模型库；通过改进的时延神经网络文本无关声纹识别算法，对样本进行相似度打分，给出用户身份预测结果，最终根据结果响应用户，并将相关数据报送后端。

后端方面，存储前端发送的声纹音频；在终端请求获取数据进行比对时，根据查询结果作出响应；对终端声纹识别结果（成功、失败、异常等）进行日志记录。

（2）**命令系统：**

具有解析用户语音对应的文本内容，将其映射至预设的简单指令集，并向测试用受控外设发出指令的功能。

后端方面，在收到终端的语音识别请求后，调用基于主流开放平台的“短语音识别”网络接口，获取预先录制的说话人音频文件所对应的文本内容，通过语义分析提取关键信息，根据预设简单指令集，映射为带有参数的简单指令，以响应的形式报送终端。

终端方面，在成功验证说话人身份并向后端发送语音识别请求后，根据后端响应中的指令操作测试用受控外设，并向用户作出指令执行结果的提示。

（3）**管理系统：**

具有管理系统配置、用户信息、声纹模型数据库、运行日志等功能。

管理系统部署于项目后端，在WEB应用中实现上述功能。管理员可登录此WEB应用进行上述操作。该系统同样可用于项目系统的监控，例如基于WebSocket或HTTP轮询方式检查在线声纹识别设备数量及状态，或根据声纹识别请求数量判断是否存在恶意使用现象，及时生成工单或进行告警，通知管理员采取相应措施。

在完成项目软硬件最终技术选型与细节设计后，项目团队展开详细设计，遵照“自顶而下，逐步求精”的程序设计思想，秉持鲁棒性、复用性、高内聚、低耦合的设计原则，对上述系统与子系统进行编码实现。

## **2.4 详细设计**

### 2.4.1 综述

### 2.4.2 用户终端应用

### 2.4.3 服务器

### 2.4.4 硬件网关

### 2.4.5 后台管理系统

# **3 项目系统测试**

## **3.1 单元测试**

## **3.2 集成测试**

# **4 产品文档**

# **5 项目总结**

## **5.1 成员分工**

## **5.2 开发日志**

# **附录 参考文献与引用**

# **附录 项目文件结构**