IIFAA

互 联 网 金 融 身 份 认 证 联 盟 标 准

T/IIFAA 3002.1—2021

远程声纹识别应用技术规范
第 1 部分：身份验证

Technical specifications for remote voiceprint recognition

Part1:Identity verification

2021-01-22 发布　　2021-01-22 实施

互联网金融身份认证联盟　　发 布

目　　次

前　言

T/IIFAA 3002—2021《远程声纹识别应用技术规范》分为以下部分：

——第1部分 身份验证；

——第2部分 身份辨识。

本文件为T/IIFAA 3002—2021的第1部分。

本文件按照GB/T 1.1—2020给出的规则起草。

本文件由互联网金融身份认证联盟提出和归口。

本文件起草单位：北京得意音通技术有限责任公司、浙江蚂蚁小微金融服务集团股份有限公司、厦门快商通科技股份有限公司、厦门天聪智能软件有限公司、银行卡检测中心、长春吉大正元信息技术股份有限公司、恒宝股份有限公司。

本文件起草人：郑方、邬晓钧、倪鸣、成舸、黄小妮、陈继东、林冠辰、李稀敏、叶志坚、洪青阳、雷文钿、杨波、王鑫、杨宏英、陈鑫、朱进丰、赵李明。

IIFAA远程声纹识别应用技术规范 第1部分：身份验证

1 范围

本文件规范了IIFAA基于声纹识别的远程身份验证的技术要求。

本文件适用于IIFAA声纹识别中涉及远程声纹身份验证技术的研发、应用及检测。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 35273—2020 信息安全技术 个人信息安全规范

GB/T 37036.1—2018 信息安全技术 移动设备生物特征识别 第1部分：通用要求

JR/T 0164—2018 移动金融基于声纹识别的安全应用技术规范

SJ/T 11380—2008 自动声纹识别（说话人识别技术规范）

3 术语和定义

下列术语和定义适用于本文件。

3.1

身份验证 identity verification

通过一定的技术手段完成对用户身份的确认。

4 缩略语

下列缩略语适用于本文件。

FAR 错误接受率（False Acceptance Rate）

FRR 错误拒绝率（False Rejection Rate）

FDR 误检率（False Detection Rate）

MDR 漏检率（Miss Detection Rate）

5 远程声纹身份验证

5.1 概述

远程声纹身份验证是指通过网络，利用以声纹识别为主的技术来验证说话人身份的方法。远程声纹身份验证实现了在无监督的非现场情况下，对说话人身份的验证，声纹识别技术的应用使得远程身份验证更加准确、安全、可靠且便捷。

5.2 整体框架

远程声纹身份验证系统架构如图1所示，其中主要包括服务器、客户端、声纹数据库。远程声纹身份验证系统首先通过客户端发起声纹身份验证请求，服务器生成验证文本发回至客户端，客户端将说话人录音发回服务器进行特征提取、身份验证和信息存储。最后，由客户端向说话人显示验证结果。

注：本文件覆盖范围为客户端与服务器、服务器与声纹数据库的通信信道两部分，主要针对1:1的验证前提进行规范。

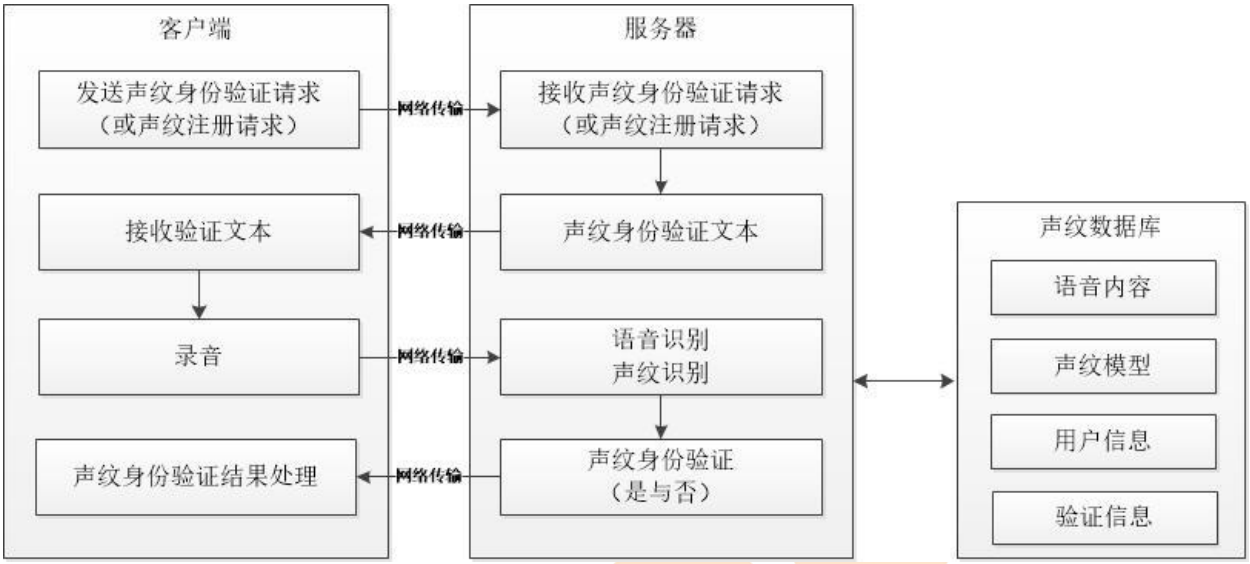

图 1　IIFAA 远程声纹身份验证架构与流程示意图

6　功能要求

6.1　声纹身份注册

远程声纹身份验证系统应具备声纹身份注册功能，包括语音信息的采集、传输、声纹模型的建立、声纹特征存储和用户信息的绑定等。

在远程声纹身份验证框架下（见图2），声纹身份注册的流程如下：

a)　注册前通过输入用户信息和账户登录的方式对用户身份进行验证，获得用户身份标识；

b)　客户端向服务器提供用户身份标识，发起声纹身份注册请求；

c)　以下操作执行 1 次或多次：

1)　服务器向客户端提供身份注册文本；

2)　客户端录制包含注册文本信息的说话人语音，并传回服务器。

d)　服务器利用所接收的语音进行声纹模型训练；

e)　服务器将训练得到的声纹模型、采集到的声纹数据与用户身份标识一起存储到声纹数据库。

在上述a）－e）中任一步骤发生系统错误，即中止声纹身份注册。

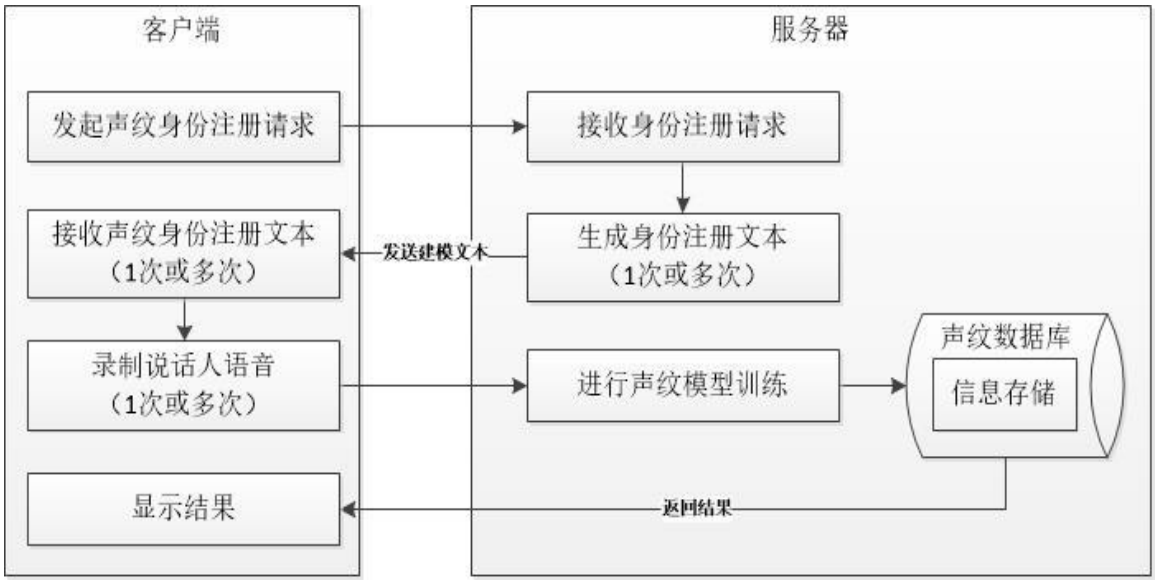

图 2　声纹身份注册流程图

6.2　声纹身份验证

远程声纹身份验证系统应具备声纹身份验证功能，包含动态声纹密码校验和声纹验证等，实现对关联用户身份的验证。

在远程声纹身份验证框架下（见图3），声纹身份验证的流程如下：

a)　客户端向服务器提供用户身份标识，发起远程身份验证请求；

b)　以下操作执行 1 次：

1)　服务器向客户端提供随机的远程身份验证文本；

2)　客户端录制说话人语音，并传回服务器。

c)　服务器利用所接收的语音，从内容和声纹特征两个方面验证客户端是否为真实的用户身份标识所代表的说话人；

d)　服务器将采集到的声纹数据存储到声纹数据库。

上述a）—d）中的任一步骤发生系统错误，即中止身份验证。

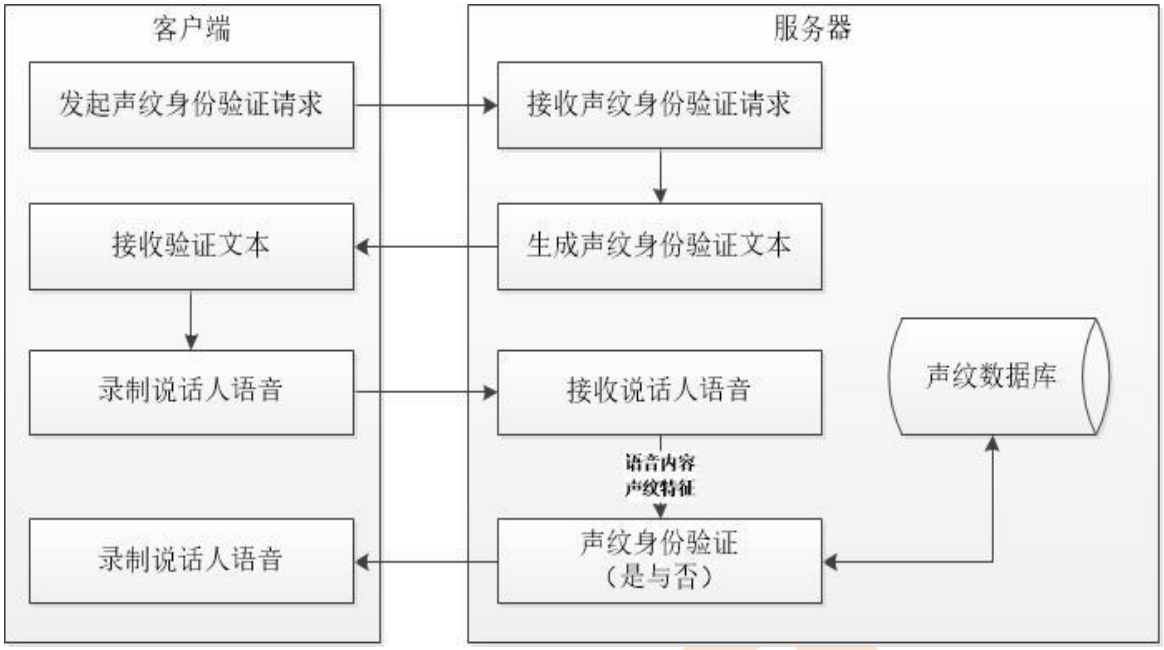

图 3　声纹身份验证流程图

6.3　声纹身份注册文本

远程声纹身份验证系统应具备提供声纹身份注册文本的功能，包括固定短语、纯数字和不限内容，具体要求如下：

a）　当声纹身份注册文本为固定短语时，文本内容应不少于 4 个汉字，且至少有 60%的汉字发音不同；

b）　当声纹身份注册文本为纯数字时，文本内容应不少于 20 个数字，且每个数字至少出现 1 次；

c）　当声纹身份注册文本不限内容时，应保证正常朗读出文本内容的语音长度不少于 10 秒。

6.4　声纹身份验证文本

远程声纹身份验证系统应具备提供声纹身份注册文本的功能，包括固定短语、纯数字和不限内容，具体要求如下：

a）　当声纹身份验证文本为固定短语时，应与注册文本相同；

b）　当声纹身份验证文本为纯数字时，验证文本内容应不多于 8 个数字；

c）　当声纹身份验证文本不限内容时，应保证正常朗读出文本内容的语音长度不少于 5 秒。

6.5　文本一次性

文本一次性要求只针对验证文本中的固定短语和纯数字，并保证在有限时间内(一般为 60s)基于已生成的动态密码，用户的单次操作有效。

当文本为不限内容时无要求。

6.6　声纹身份更新

远程声纹身份验证系统应具备声纹身份更新功能。声纹身份更新，即更新用户身份标识所对应的声纹模型。在远程声纹身份验证框架下，声纹身份更新有以下几种情况：

a) 服务器端根据预先定义的声纹身份更新机制，利用声纹数据库中存储的声纹数据，自动进行声纹模型的更新；
b) 因具体应用需要、技术升级等原因，服务器端利用声纹数据库中存储的声纹数据，重新进行声纹模型训练；
c) 用户主动发起声纹身份更新请求，在验证用户真实身份后，重新进行声纹身份注册。

6.7 声纹身份注销

远程声纹身份验证系统应具备声纹身份注销功能，具体要求如下：
a) 因具体应用需要，服务器端可在声纹数据库中删除用户身份标识及相应的声纹模型和声纹数据，从而注销相应用户的声纹身份信息；
b) 用户主动发起声纹身份注销请求，在验证用户真实身份后，服务器端注销相应用户的声纹身份信息；
c) 声纹身份注销后，应确保不可被检索、访问。

7 性能要求

7.1 响应时间

7.1.1 声纹身份注册时间

从服务器完全接收到所有用于声纹身份注册的语音开始，到完成相应声纹身份注册，所需时间应小于等于1秒。

7.1.2 声纹身份验证时间

从服务器完全接收到所有用于声纹身份验证的语音开始，到完成相应声纹身份验证，所需时间应小于等于0.5秒。

7.2 性能评价指标

评价远程声纹验证的性能，使用如下指标：
a) 错误接受率（FAR）

$$\text{错误接受率(FAR)} = \frac{\text{非目标语音判断为目标的结果数}}{\text{冒充测试次数}} \times 100\%$$

b) 错误拒绝率（FRR）

$$\text{错误拒绝率(FRR)} = \frac{\text{目标语音判断为非目标的结果数}}{\text{目标测试次数}} \times 100\%$$

7.3 声纹身份验证的精度与分级

声纹身份验证精度分为一般级和增强级。
一般级：FAR≤1.0%时，FRR≤2.5%；
增强级：FAR≤0.5%时，FRR≤3.0%。

7.4 控制声纹时变的影响

声纹随时间推移会发生自然的变化，导致真实用户的声纹身份验证错误拒绝率升高。由于声纹时变的影响，导致用户声纹身份验证的错误拒绝率高于以下指标要求时，应进行声纹身份的重新注册：

一般级：FRR=5.0%；

增强级：FRR=6.0%。

7.5 增强的远程声纹身份验证性能要求

7.5.1 减小跨信道的影响

声纹在不同客户端设备切换的时候，例如不同的采集设备，不同的传输信道之间的跨信道会导致真实用户的声纹身份验证错误拒绝率升高。利用针对性的声纹识别算法，减小设备间的跨信道对身份验证正确率的影响。

8 安全要求

8.1 概述

远程声纹身份验证系统应符合GB/T 37036.1—2018、GB/T 35273—2020中的规定，在此基础上还应满足以下安全要求。

8.2 数据存储

远程声纹身份验证涉及到的声纹信息存储应满足以下要求：

a) 声纹模型和声纹数据只允许保存在服务器上，不应临时或长时保存在客户端。

b) 服务器端应加密保存声纹身份注册、声纹身份验证流程中有关的信息，采取高强度安全防护措施防止未授权访问、泄露、篡改或者毁损。

8.3 网络通信

服务器与客户端间的网络通信应采用安全传输协议。

8.4 管理要求

应符合如下要求：

a) 应具备有效的安全机制，确保当前操作人员拥有合法权限完成用户登记、更新和注销；

b) 宜采取适当的机制和程序，在用户登记过程中确认当前登记者的真实身份，确保登记的用户声纹数据与该用户的身份标识之间的正确关联；

c) 若声纹识别支持不同用户使用权限，应具备有效的安全机制确保不同权限用户只能在其授权范围内进行相应操作。

8.5 安全环境

基于GB/T 37036.1—2018中对安全环境的规定，若移动设备支持可信执行环境或安全单元等安全环境，在声纹数据采集、存储和比对过程中：

a) 宜使用位于可信执行环境中的声纹数据采集模块对用户的声纹数据进行采集；

b) 宜在可信执行环境中对采集的用户声纹数据进行质量判断、呈现攻击检测；

c) 应通过可信执行环境中可信交互界面实现与用户之间的交互；

d) 如需与位于富执行环境声纹数据采集模块或移动应用进行数据交互时，应具备有效的安全机制验证富执行环境中交互对象的合法性，数据交互过程中宜采用安全通道机制以保证交互数据的完整性和机密性。

8.6 日志安全要求

日志安全要求包括但不限于：

a) 日志记录中不应出现明文的声纹数据、密钥信息或其他安全相关的参数等；

b) 应具备有效的安全机制，防止对日志记录的非授权访问；

c) 应具备授权管理机制，对日志记录的写入、读取、删除的操作权限进行管理。

8.7 增强的远程声纹身份验证安全性

8.7.1 概述

服务器端可通过技术手段实现防止录音重放假冒、防止拼接假冒以及防止转换或合成语音功能，以增强远程声纹身份验证的正确性和安全性。

8.7.2 防止录音重放假冒

对接收到的语音进行检测，防止攻击者采用高保真录音设备录制或通过其他手段获取用户语音样本来试图通过身份验证。

8.7.3 防止拼接假冒

对接收到的语音进行检测，防止攻击者将用户身份标识对应者的语音进行拼接并试图通过身份验证。

8.7.4 防止转换或合成语音

对接收到的语音进行检测，防止攻击者采用语音合成技术，生成用户身份标识对应者的语音并试图通过身份验证。

8.8 防攻击检测性能

评价远程声纹验证的防攻击检测性能，使用如下指标：

a) 误检率（FDR）

$$\text{误检率(FDR)} = \frac{\text{正常语音判断为攻击语音个数}}{\text{正常语音个数}} \times 100\%$$

误检率的数值越低，系统的防攻击检测性能越好。

远程声纹验证的误检率应＜3.0%。

b) 漏检率（MDR）

$$\text{漏检率(MDR)} = \frac{\text{攻击语音判断为正常语音个数}}{\text{攻击语音个数}} \times 100\%$$

漏检率的数值越低，系统的防攻击检测性能越好。

远程声纹验证的漏检率应＜0.5%。

附 录 A
（资料性附录）
精度等级适用的应用场景分类

表 A.1 精度等级适用的应用场景

序号	精度等级	应用场景
1	一般级	用户登录
2	增强级	资金交易类、用户信息变更、业务变更等