ICS 03.060
A11

JR

中华人民共和国金融行业标准

JR/T 0164—2018

移动金融基于声纹识别的安全应用技术规范

Technical specifications for voiceprint recognition based security application for mobile finance

2018－10－09 发布　　2018－10－09 实施

中国人民银行　　发布

目　次

前　言

本标准依据GB/T 1.1—2009给出的规则起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本标准由全国金融标准化技术委员会（SAC/TC 180）归口。

本标准主要起草单位：中国建设银行股份有限公司、清华大学、中国工商银行股份有限公司、中国农业银行股份有限公司、中国银行股份有限公司、北京得意音通技术有限责任公司、北京银联金卡科技有限公司、交通银行股份有限公司、中信银行股份有限公司、中国光大银行股份有限公司、华夏银行股份有限公司、中国民生银行股份有限公司、招商银行股份有限公司、兴业银行股份有限公司、广发银行股份有限公司、平安银行股份有限公司、上海浦东发展银行股份有限公司、恒丰银行股份有限公司、浙商银行股份有限公司、渤海银行股份有限公司、中国邮政储蓄银行股份有限公司、中国银联股份有限公司、中国金融电子化公司、北京软件产品质量检测检验中心、中金金融认证中心有限公司、信息产业信息安全测评中心、浙江蚂蚁小微金融服务集团股份有限公司、财付通支付科技有限公司、百度在线网络技术（北京）有限公司、北京京东金融科技控股有限公司、第四范式（北京）技术有限公司、同盾科技（杭州）有限公司、北京眼神智能科技有限公司、科大讯飞股份有限公司。

本标准主要起草人：金磐石、郑方、刘建忠、郭汉利、张晓东、杨杰、邓玉、邓立峰、张玉、许剑锋、廖敏飞、刘芳、刘红波、刘丽娟、许腾、邬晓钧、李蓝天、李通旭、米青山、黄维登、王宁、孙毅、程尧、杨晔萌、肖永明、成舸、倪鸣、陈柳村、王小钢、刘乐、曹慧、渠韶光、孟飞宇、高志民、高强裔、于柳绮、白阳、陈聪、朱京城、陶宏、郭勇、陈海、裴云龙、柯技、陈云峰、吴永飞、陈刚、古伟、虞刚、聂建军、黄艺驰、李金堆、陈桂斌、赖众程、张亮、李健、杨耀勇、张爽、李顺达、臧铖、王晓琳、陈震宇、江黎枫、谭颖、汪之婴、黄梦达、谢荣东、安荔、邬向阳、张明哲、倪又明、张锡铭、王秀君、马洪涛、王飞宇、高峰、陈硌琼、于璐、刘健、王冠华、陈星、落红卫、宋铮、吴永强、张翔、苏振东、耿琦、冯璐、高卓、青飞、王一鹤、曾谁飞、杨春林、马万钟、蒯天祥。

移动金融基于声纹识别的安全应用技术规范

1 范围

本标准规定了移动金融服务场景中基于声纹识别的安全应用的功能要求、性能要求和安全要求等内容，不包括电话或网络电话（VoIP）中涉及声纹识别的应用场景。

本标准适用于移动金融服务基于声纹识别的设计、开发、检测、应用及风控。

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。

JR/T 0118 金融电子认证规范

GM/T 0021 动态口令密码应用技术规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

移动设备标识 mobile device identity

国际移动设备识别码（IMEI）、电子序列号（ESN）、移动设备识别码（MEID）、广告识别符（IDFA）、国际移动用户识别码（IMSI）等能唯一标识移动设备的编码。

3.2

声纹 voiceprint

对语音中所蕴含的、能表征和标识说话人的语音特征，以及基于这些特征（参数）所建立的语音模型的总称。

3.3

声纹特征 voiceprint feature

从说话人的语音中所提取出来的、可以表征该说话人语音的个性特征的参数。

注：常用的特征参数包括频谱(spectrum)、倒频谱(cepstrum)、线性预测系数(LPC)、音高(pitch)、声调(tone)、共振峰(formant)、音质(voice quality)、声韵(prosody)、习语(phoneme/word idiolect)等各种层次的信息。

3.4

声纹模型 voiceprint model

对声纹特征进行描述的数学模型。

注：常用的数学模型有：高斯混合模型(Gaussian mixture model)，基于通用背景模型的高斯混合模型(Gaussian

mixture model-universal background model)，隐马尔可夫模型(hidden Markov model)，人工神经网络(artificial neural network)，支持向量机(support vector machine)等。

3.5

声纹模型训练 voiceprint model training

从说话人的有效语音提取声纹特征并根据声纹特征估计其声纹模型的参数的过程。

3.6

声纹识别 voiceprint recognition(VPR)

根据待识别语音的声纹特征识别该段语音所对应的说话人的过程。

注：声纹识别包含声纹确认和声纹辨认，在本标准中涉及的声纹识别特指声纹确认。

3.7

声纹辨认 voiceprint identification

给定一段语音和一组候选说话人的声纹模型，判断该段语音是哪位说话人所说的声纹识别方式。

注：声纹辨认是一个“多选一”的问题。

3.8

声纹确认 voiceprint verification

给定一段只含一名说话人的语音和一个说话人的声纹模型,判断该段语音是否是该说话人所说的声纹识别方式。

注：该段语音通常也称为“待识别”语音，该说话人通常也称为“宣称的说话人”。声纹确认系统的输出是一个“二值判别”，它的结果只有两种：接受或拒识。

3.9

接受 acceptance

声纹识别系统判定待识别语音是宣称说话人所说。

3.10

拒识 rejection

声纹识别系统判定待识别语音不是宣称说话人所说。

3.11

错误接受 false acceptance

声纹识别系统将非宣称说话人的语音判断为宣称说话人的语音。

3.12

错误接受率 false acceptance rate(FAR)

声纹识别过程中错误接受的数目占测试集合中应被拒绝的测试数目的百分率。

3.13

错误拒绝 false rejection

声纹识别系统将宣称说话人的语音错误地判断为非宣称说话人的语音。

3.14

错误拒绝率 false rejection rate(FRR)

声纹识别过程中错误拒绝的数目占测试集合中应被接受的测试数目的百分率。

3.15

动态声纹密码 dynamic voiceprint code

对用户进行身份验证时，系统基于服务端随机（或基于服务端安全算法）产生动态文本，如数字串、字母串或它们组合的串等。

3.16

录音欺诈 spoofing by replay

在声纹确认过程中，通过播放已经录制好的目标用户的声音尝试通过声纹验证的行为。

3.17

录音拼接欺诈 spoofing by replay of connected speech segments

在声纹确认过程中，把已经录制好的目标用户录音片段通过软件拼接成待验证语音，然后播放尝试通过声纹验证的行为。

3.18

有效语音 valid speech

进行声纹模型训练或声纹识别的语音中，抛除静音、背景噪音等不含有说话人信息的无效语音后的语音。

3.19

声纹信息控制者 voiceprint data controller

有权决定声纹信息处理目的、方式等的组织或个人。

注：本标准中所指声纹信息包含声纹模型信息和语音信息。

3.20

明示同意 explicit consent

用户通过书面声明或主动做出肯定性动作，对其个人信息进行特定处理做出明确授权的行为。

注：肯定性动作包括用户主动作出声明（电子或纸质形式）、主动勾选、主动点击“同意”、“注册”等。

3.21

匿名化 anonymization

通过对个人信息的技术处理，使得用户无法被识别，且处理后的信息不能被复原的过程。

4 应用概述

4.1 声纹识别应用

声纹识别是根据待识别语音的声纹特征鉴别该段语音所对应的说话人的过程，在移动金融服务中可用于用户的身份认证。

4.2 声纹识别应用流程

在移动金融服务中基于声纹识别的应用流程如图1所示，用户通过拾音设备进行语音采集，经移动金融客户端加密传输至服务器端。客户端前置服务器进行必要的业务处理后将语音信息传输至声纹服务器。声纹服务器完成声纹的注册、验证、变更或注销，并将相应的结果（接受或拒识）经客户端前置服务器反馈至移动金融客户端。

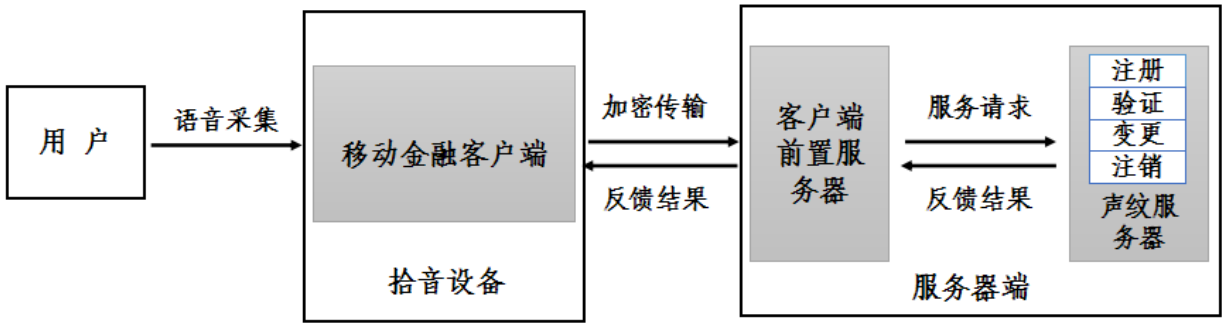

图1 声纹识别应用流程示意图

5 功能要求

5.1 声纹注册

声纹的注册功能，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身份的绑定等。

声纹注册要求如下：

——注册前需对用户身份进行认证。

——注册前应取得用户的明示同意。

——传输时应包含用户属性数据，如用户唯一性标识、移动设备标识等。

——存储时应与用户属性数据形成映射关系。

5.2 声纹验证

声纹的验证功能，包含动态声纹密码校验和声纹确认等，实现对关联账户主体身份的验证。

动态声纹密码应由服务器端生成，且要求如下：

——长度宜 6 位或 8 位。

——有效期不超过 120 秒。

——避免连续重复，如“…11…”等。

——验证后应及时清除。

5.3 声纹变更

声纹的变更功能，即服务器端重新进行声纹模型训练以实现对原有声纹信息的更新。

声纹变更根据发起方不同分为系统主动发起、声纹信息控制者主动发起和用户主动发起三种：

——系统主动发起的变更是根据系统定义的模型更新机制对声纹模型进行变更。

——声纹信息控制者主动发起的变更是因业务、技术等因素需变更算法或调整模型架构，如算法厂

商变更、算法或模型重设计等。

——用户主动发起的变更是由用户主动意愿发起的自有声纹模型的变更;用户主动发起的声纹变更应满足以下要求：

- 变更前应对用户身份进行认证。
- 变更前应取得用户的明示同意。

5.4 声纹注销

声纹注销根据发起方不同分为声纹信息控制者主动发起和用户主动发起两种：

——声纹信息控制者主动发起对声纹的注销功能时，应满足以下要求：

- 明确告知用户注销原因。
- 明确告知用户注销时间。

——用户主动发起对声纹的注销功能应满足以下要求：

- 注销前应对用户身份进行认证。
- 注销前应取得用户的明示同意。

声纹注销后应删除与用户相关的声纹信息或做匿名化处理，不应重复使用。

6 性能要求

6.1 基本性能指标

基本性能指标应满足以下要求：

——错误接受率(FAR)≤0.5%。

——错误拒绝率(FRR)≤3.0%。

6.2 采样指标

采样指标应满足以下要求：

——采样率：16kHz。

——采样精度：16bit。

6.3 有效语音长度

有效语音长度应满足以下要求：

——声纹注册时：有效语音长度≥5000ms。

——声纹验证时：有效语音长度≥1000ms。

6.4 系统响应时间

系统响应时间应满足以下要求：

——声纹的注册时：系统响应时间≤3000ms。

——声纹的验证时：系统响应时间≤2000ms。

6.5 语音信息质量判断

应具有语音信息质量判断的能力，包括但不限于截幅比例、信噪比、完整程度。

6.6 抗噪音能力

应具有抗噪音能力，以保证系统的可用性。

6.7 抗时变能力

应具有因时间变化而导致声音变化的正确处理能力，以保证系统的可用性。

7 安全要求

7.1 声纹信息采集

7.1.1 基本要求

声纹信息采集基本要求如下：

——应使用动态声纹密码。

——采集完成后，应立即对声纹信息进行加密处理。

——应采取安全措施保证声纹信息不被其他设备或程序非授权获取。

——应采取防篡改机制保证声纹信息不被其他设备或程序篡改。

——声纹注册时，声纹信息采集宜使用多组动态声纹密码。

7.1.2 身份认证要求

声纹注册采集语音信息前，应使用多种要素验证用户身份，采用以下方式之一：

——采用符合 JR/T 0118 的数字证书，并组合交易密码等至少一种认证要素。

——采用符合 GM/T 0021 的动态令牌设备，并组合交易密码等至少一种认证要素。

——至少组合两种认证要素（其中至少一种为动态认证要素，如动态验证码、基于客户行为的动态挑战应答等），并采用短信、数据（如手机银行、即时通讯、邮件）等至少两种不同通信渠道。

7.1.3 明示同意要求

应向被采集用户进行明示，明确告知声纹信息收集、使用信息的目的、方式和范围，征得用户同意后方可进行采集。

7.2 声纹信息传输

声纹信息传输应满足以下要求：

——应采用安全传输协议，保证声纹信息传输时的完整性和保密性。

——声纹识别接口仅限于服务器端之间内部调用，不应暴露在公共、开放的网络上。

7.3 声纹信息存储

声纹信息存储应满足以下要求：

——客户端应用软件应禁止以任何形式留存声纹信息，包含但不限于声纹采集、声纹验证等过程中使用的声纹信息。

——服务器端应加密保存声纹模型，并防止声纹模型的未授权访问、泄露、篡改或者毁损。

——服务器端如留存语音信息，应对语音信息进行加密或采取高强度安全防护措施防止语音信息的未授权访问、泄露、篡改或者毁损；应对语音信息去标识化或脱敏处理，以确保对外不可用。

——服务器端留存的声纹模型信息保存时间应为实现目的所必需的最短时间。

——在发生或者可能发生声纹信息遗失、泄露或者毁损等情况时，应当立即采取补救措施，及时告知用户。

——在境内运营中采集和产生的声纹信息应当在境内存储。因业务需要，确需向境外提供的，应符合相关的法律法规要求。

7.4 声纹信息处理

7.4.1 基本要求

声纹信息处理应满足以下基本要求：

——应采取有效措施，防止声纹模型配置参数的未授权访问、泄露、篡改等。

——在声纹注册、验证、变更和注销各个环节，应对关键操作信息进行日志记录。

——用户主动发起声纹变更、注销前，应采用 7.1.2 中要求的身份认证方式验证用户身份。

——应具有失败处理措施，在失败时进行相应提示并限制失败次数，如果超过限制次数，应触发相应的失败控制机制。

——声纹信息不应转让，禁止用于声纹注册、验证、变更、注销之外的其他用途，法律另有规定的除外。

——不得向其他客户端应用软件提供声纹信息。

7.4.2 防攻击能力

应具备抵御常见攻击的能力，包括但不限于：

——防语音模仿：在声纹确认过程中，应能够抵御攻击者模仿说话人、试图以说话人的身份通过声纹验证的攻击行为。

——防语音转换及合成：在声纹确认过程中，应能够抵御攻击者通过机械的、电子的方法产生人造语音的攻击行为，如语音合成技术。

——防录音欺诈：在声纹确认过程中，应能够抵御播放已录制好的目标用户声音并尝试通过声纹验证的攻击行为。

——防录音拼接欺诈：在声纹确认过程中，应能够抵御将已录制好的目标用户录音片段拼接成待验证语音播放并尝试通过声纹验证的攻击行为。

7.5 声纹信息删除

声纹信息删除后，应确保不可被检索、访问。

参 考 文 献

[1] 中华人民共和国网络安全法（全国人民代表大会常务委员会 2016年11月7日发布，2017年6月1日实施）

[2] GB/T 25069—2010 信息安全技术 术语

[3] GB/T 35273—2017 信息安全技术 个人信息安全规范

[4] GA/T 893—2010 安防生物特征识别应用术语

[5] GA/T 1179—2014 安防声纹确认应用算法技术要求和测试方法

[6] JR/T 0068—2012 网上银行系统信息安全通用规范

[7] JR/T 0071—2012 金融行业信息系统信息安全等级保护实施指引

[8] JR/T 0092—2012 中国金融移动支付 客户端技术规范

[9] SJ/T 11380—2008 自动声纹识别（说话人识别）技术规范