ICS 35.240.01
CCS L 67

中华人民共和国国家标准

GB/T 37036.5—2023

信息技术　移动设备生物特征识别 第5部分：声纹

Information technology—Biometrics used with mobile devices—Part 5：Voiceprint

2023-03-17 发布　　2023-10-01 实施

国家市场监督管理总局
国家标准化管理委员会　发布

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

本文件是GB/T 37036《信息技术　移动设备生物特征识别》的第5部分。GB/T 37036已经发布了以下部分：

——第1部分：通用要求；

——第2部分：指纹；

——第3部分：人脸；

——第4部分：虹膜；

——第5部分：声纹；

——第6部分：指静脉。

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专利的责任。

本文件由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本文件起草单位：北京得意音通技术有限责任公司、中国电子技术标准化研究院、清华大学、北京曙光易通技术有限公司、厦门快商通科技股份有限公司、北京眼神科技有限公司、蚂蚁科技集团股份有限公司、华为技术有限公司、科大讯飞股份有限公司、北京百度网讯科技有限公司、福州数据技术研究院有限公司、中国人民大学、浙江鼎立实业有限公司、哈尔滨理工大学。

本文件主要起草人：郑方、王文峰、钟陈、邬晓钧、黄小妮、于雪平、李稀敏、宋继伟、刘倩颖、杨春林、林冠辰、孟凡辉、马万钟、李旭、陈福泓、金琴、俞云祥、陈晨、杨兴亚、叶志坚、彭晋。

引　言

GB/T 37036《信息技术　移动设备生物特征识别》拟由九个部分构成。

——第1部分：通用要求。目的在于确立移动设备生物特征识别的技术架构、通用流程和要求。

——第2部分：指纹。目的在于确立移动设备在指纹生物识别方面的流程和要求。

——第3部分：人脸。目的在于确立移动设备在人脸生物识别方面的流程和要求。

——第4部分：虹膜。目的在于确立移动设备在虹膜生物识别方面的流程和要求。

——第5部分：声纹。目的在于确立移动设备在声纹生物识别方面的流程和要求。

——第6部分：指静脉。目的在于确立移动设备在指静脉生物识别方面的流程和要求。

——第7部分：多模态融合。目的在于确立移动设备在多模态生物识别方面的流程和要求。

——第8部分：呈现攻击检测。目的在于规定移动设备呈现攻击检测要求。

——第9部分：测试方法。目的在于描述移动设备不同类型生物特征识别的功能和性能测试方法。

《信息技术　移动设备生物特征识别》标准第1部分给出了总体架构，提出了通用的要求；后续部分根据生物特征识别技术在移动设备上的发展情况，针对指纹、人脸、虹膜、声纹、指静脉等具体模态提出具体要求，最后从安全和性能出发，对呈现攻击检测以及性能测试做出了规范，指导产品的落地与评估。

信息技术　移动设备生物特征识别
第5部分：声纹

1　范围

本文件描述了移动设备声纹识别的技术架构，规定了移动设备声纹识别的功能要求、性能要求和安全要求。

本文件适用于移动设备声纹识别产品的设计、生产、集成和应用。

2　规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；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 37036.1—2018　信息技术　移动设备生物特征识别　第1部分：通用要求

GB/T 40660—2021　信息安全技术　生物特征识别信息保护基本要求

ISO/IEC 30107-3　信息技术　生物特征识别呈现攻击检测　第3部分：测试与报告(Information technology—Biometric presentation attack detection—Part 3：Testing and reporting)

3　术语和定义

下列术语和定义适用于本文件。

3.1

声纹语音样本　voiceprint speech sample

可提取声纹的语音样本。

注1：智能语音交互过程中所收集的语音样本如未经过特殊处理，可提取声纹，属于声纹语音样本。

注2：采用参数合成方法生成的语音样本不包含声纹，不属于声纹语音样本。

注3：声纹语音样本是一类生物特征样本，生物特征样本见GB/T 5271.37。

3.2

声纹数据　voiceprint data

声纹语音样本及其处理得到的，可单独或结合其他信息识别数据主体的数据。

注：声纹数据包括声纹语音样本、声纹特征项和声纹模型，也称为声纹识别数据。

3.3

声纹注册　voiceprint enrolment

根据策略创建和存储声纹模型数据记录的行为。

[来源：GB/T 5271.37—2021，3.5.3，有修改]

3.4

声纹建模　voiceprint modeling

根据说话人的声纹特征项估计其声纹模型参数的过程。

注：声纹建模也可称为声纹模型训练。

4 缩略语

下列缩略语适用于本文件。

DVPC 动态声纹密码(Dynamic Voiceprint Code)

FAR 错误接受率(False Acceptance Rate)

FRR 错误拒绝率(False Rejection Rate)

5 技术架构

本文件给出的移动设备声纹识别技术架构是 GB/T 37036.1—2018 中描述的通用技术架构在声纹应用领域的具体化,见图 1。

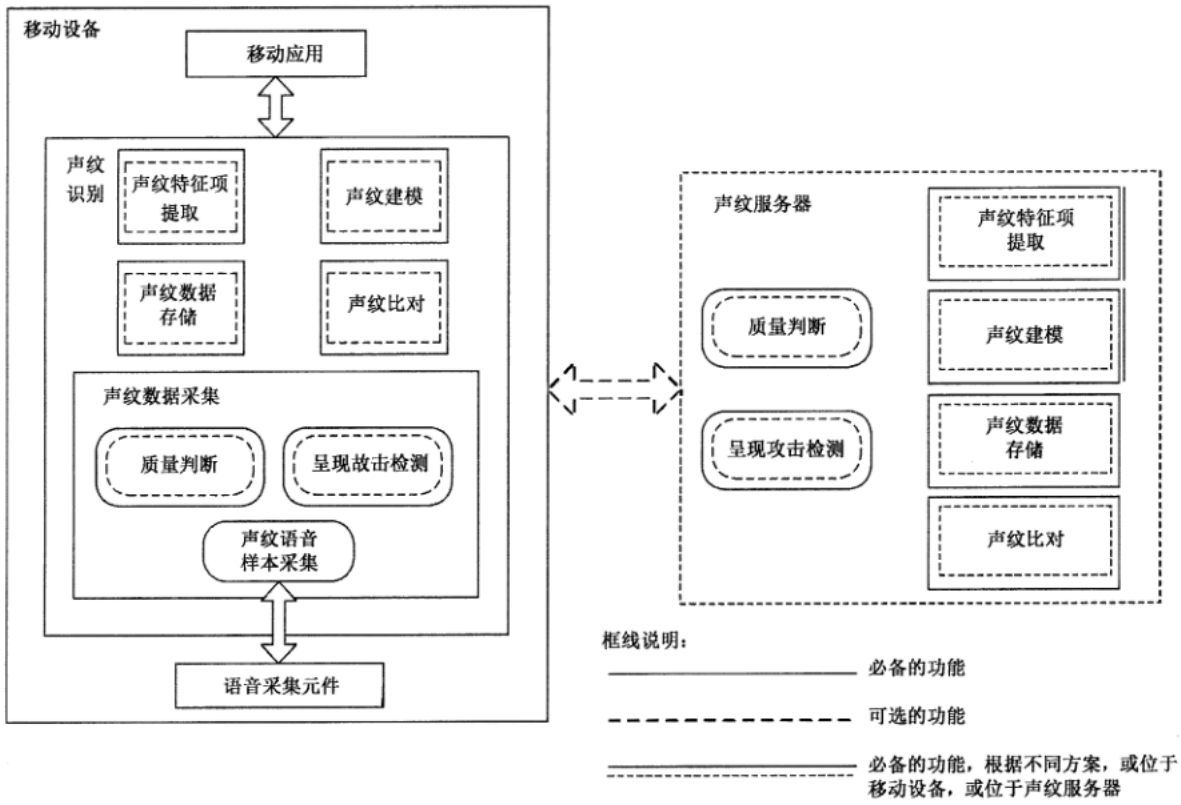

图 1 移动设备声纹识别技术架构

移动设备上声纹识别由语音采集元件、声纹识别、移动应用三部分组成。其中,语音采集元件提供声纹语音样本采集功能。声纹识别提供声纹数据采集、声纹特征项提取、声纹建模、声纹数据存储、声纹比对功能。移动应用是移动设备中的声纹识别的服务调用方,可为一个独立的移动应用软件、移动应用软件中的一个功能模块或移动设备操作系统提供的一个系统服务。移动设备声纹识别有两种应用模式,分别为位于移动设备中的本地识别模式(即非联网模式)和结合声纹服务器的远程识别模式(即联网模式)。

在本地识别模式中,声纹语音样本采集、声纹特征项提取、声纹建模、声纹数据存储和声纹比对均在移动设备中实现,并在声纹语音样本采集过程中进行质量判断和呈现攻击检测。

在远程识别模式中,声纹语音样本采集在移动设备中实现,通过调用移动设备中的语音采集元件采

集用户的语音，用于后续的声纹特征项提取、声纹建模、声纹数据存储和声纹比对；质量判断、呈现攻击检测和声纹特征项提取可在移动设备中实现，也可以在声纹服务器端实现；声纹建模、声纹数据存储和声纹比对一般在服务器端中实现。

一些典型应用模式示例见附录A。

6 业务流程

移动设备声纹识别业务主要流程包括声纹注册过程、声纹识别过程和声纹注销过程。

a) 声纹注册过程，包括但不限于如下步骤。
   1) 对用户进行身份鉴别和权限检测，如果验证通过则进行下一步，否则不准许进行声纹注册；
   2) 启动声纹注册过程；
   3) 进行声纹语音样本采集；
   4) 对采集的声纹语音样本进行质量判断、呈现攻击检测，如果未通过检测则重新进行声纹语音样本采集；
   5) 对通过检测的声纹语音样本进行声纹特征项提取，并根据所提取的特征项计算出声纹模型进行存储，并与用户身份标识进行关联；
   6) 结束声纹注册过程。
b) 声纹识别过程，包括但不限于如下步骤：
   1) 启动声纹识别过程；
   2) 进行声纹语音样本采集；
   3) 对采集的声纹语音样本进行质量判断、呈现攻击检测，如果检测通过则进行下一步，否则退出识别过程；
   4) 提取用户声纹语音样本中的声纹特征项，直接或进行声纹建模后与存储的一个或多个用户声纹模型进行比对；
   5) 根据比对结果做出判断决策，并传输至移动应用；
   6) 结束声纹识别过程。
c) 声纹注销过程，包括但不限于如下步骤：
   1) 启动声纹注销过程；
   2) 删除与待注销用户关联的声纹数据以及身份标识；
   3) 结束声纹注销过程。

7 功能要求

7.1 一般要求

7.1.1 基本功能

符合GB/T 37036.1—2018中6.1.1的基本功能要求，包括但不限于：

a) 应适用于不同人种、不同年龄可正常说话的用户；
b) 应适用于移动设备用户和声纹识别系统管理员；
c) 应支持对声纹比对阈值的设定；
d) 宜可结合其他生物特征识别技术实现多模态融合识别。

7.1.2 功能管理

符合 GB/T 37036.1—2018 中 6.1.2 的功能管理要求，包括但不限于：

a) 应支持新用户注册，已注册用户声纹模型删除、更新、注销等功能；

b) 应支持用户注册并存储声纹模型；

c) 宜支持用户、声纹识别系统管理员等不同用户使用权限，在声纹识别中的声纹语音样本采集、声纹数据存储和声纹比对等功能中分别具有相应的权限管理机制；

d) 应具备异常情况处理能力。

注：异常情况包处理能力包括声纹语音样本采集失败、声纹语音样本未通过质量检测、声纹识别决策失败、声纹更新失败、提示用户重新采集或提示失败等。

7.1.3 日志管理

移动设备声纹识别的日志管理功能应记录以下内容：

a) 声纹注册过程中的成功或失败事件及发生时间和操作人员；

b) 声纹识别过程中的成功或失败事件及发生时间和操作人员；

c) 声纹模型更新过程中的成功或失败事件及发生时间和操作人员；

d) 声纹注销过程中的成功或失败事件及发生时间和操作人员。

7.2 声纹数据采集

7.2.1 声纹语音样本采集

符合 GB/T 37036.1—2018 中 6.2.1 的要求，包括但不限于：

a) 在以声纹建模为目的的声纹语音样本采集前，结合其他身份鉴别手段对用户身份进行认证；

b) 提示用户采集的环境要尽量保持安静，避免周围环境噪声和其他说话声音的干扰，避免回音，避免录音现场的各类电磁干扰；

c) 具有明显的用户提示，告知用户对其声纹语音样本进行了采集，若采集过程分多次进行，宜向用户明示每一次采集的进度；

d) 具备异常情况判定及处理能力，如声纹语音样本未通过质量判断、检测到呈现攻击等的相应处理机制。

7.2.2 质量判断

移动设备声纹识别应具备对采集到的用户声纹语音样本进行质量判断的能力，以确定当前声纹语音样本是否满足声纹识别处理的需求，包括但不限于对截幅比例、信噪比和声纹语音样本完整程度进行质量判断。

7.2.3 呈现攻击检测

移动设备声纹识别应具备呈现攻击检测功能，且应符合 GB/T 37036.1—2018 中 6.2.3 的要求与 ISO/IEC 30107-3 的相关规定。

移动设备声纹识别的呈现攻击检测功能，应支持对下述呈现攻击类型的检测，如表 1 所示，一些可用的呈现攻击检测方法见附录 B。

移动设备声纹识别检测出呈现攻击时应具备相应的处理机制，如失败/错误提示或进行风险提示等。

表 1　声纹识别呈现攻击类型

呈现攻击类型	样例
语音模仿	攻击者模仿说话人
语音转换及合成	攻击者通过机械的、电子的方法产生人造语音
录音重放	播放已录制的目标用户语音
录音拼接	将已录制的目标用户录音片段拼接成语音播放

7.3　声纹特征项提取

移动设备声纹识别应能从通过了质量判断的用户声纹语音样本中提取用户的声纹特征项，提取过程宜采用不可逆的方式。

7.4　声纹建模

移动设备声纹识别应具备从说话人有效语音中提取声纹特征项后估计其声纹模型参数并生成数学模型的能力。

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。

7.5　声纹数据存储

符合 GB/T 37036.1—2018 中 6.3 的要求，包括但不限于：

a)　应具备声纹数据存储管理功能；
b)　应只允许具有合法权限的操作人员录入、访问、读取或删除存储的用户声纹数据；
c)　应能够把注册的用户声纹模型与该用户的身份标识进行关联；
d)　应支持同一用户根据系统提示的不同内容形式(包括文本相关、文本无关、数字串等)注册声纹数据进行存储；
e)　应具备异常情况判定及处理能力，如声纹模型存储、读取或删除失败时的相应处理机制；
f)　本地识别模式中，不应存储用户原始的声纹语音样本；
g)　远程识别模式中，需要对用户原始的声纹语音样本进行存储时，应与声纹特征项、声纹模型分开存储。

7.6　声纹比对

7.6.1　基本要求

移动设备声纹识别应提供下述至少一种识别功能：

a)　用户验证，即一对一比对；
b)　用户辨识，即一对多比对。

7.6.2　比对判定及处理

移动设备声纹识别比对判定及处理应提供以下功能：

a)　将提取的声纹特征项或生成的声纹模型和已存储的声纹模型进行比对，计算出比对得分；
b)　根据比对得分进行识别结果判定，输出识别结果；

c) 对异常情况进行判定并做出相应处理。

注：异常情况包括比对失败、识别决策失败等。

8 性能要求

8.1 基本性能指标

8.1.1 本地声纹验证指标

在本地识别模式中，用户声纹验证的基本性能指标应满足以下要求：

a) 一般级：FAR≤4.0%时，FRR≤1.5%；

b) 增强级：FAR≤1.5%时，FRR≤3.0%。

8.1.2 本地声纹辨识指标

在本地识别模式中，当集合人数小于或等于10时，用户声纹辨识的基本性能指标应满足以下要求：

a) 开集辨识：应满足本地声纹验证的基本性能指标；

b) 闭集辨识：正确识别率不低于99.0%。

8.2 采样指标

声纹语音样本采集的采样指标应满足以下要求：

a) 采样率：16 000 Hz；

b) 采样精度：16 bit。

8.3 有效语音长度

声纹语音样本的有效语音长度应满足以下要求：

a) 文本相关：声纹注册时有效语音长度大于或等于3 s，声纹识别时有效语音长度大于或等于1 s；

b) 文本无关：声纹注册时有效语音长度大于或等于10 s，声纹识别时有效语音长度大于或等于3 s；

c) DVPC：声纹注册时有效语音长度大于或等于5 s，声纹识别时有效语音长度大于或等于1 s。

8.4 质量要求

声纹语音样本质量应满足以下要求：

a) 截幅比例小于或等于10%；

注：截幅比例又称为截幅率，是指一段完整的声纹语音样本中因音量过大或过小、环境噪声等问题，导致出现语音信号振幅超出系统量化范围的现象，语音截幅时长占语音总有效时长的比例。

b) 信噪比大于或等于20 dB；

c) 声纹语音样本内容与要求的内容一致。

8.5 抗噪声能力

不同架构模式的移动设备声纹识别，在环境噪声等效声级≤50 dB(A)时，应满足本文件8.1中一般级的性能指标；在环境噪声等效声级大于50 dB(A)且小于或等于60 dB(A)时，在满足本文件8.1中一般级性能指标中的FAR条件下，FRR不超过限值的2倍。

注：dB(A)的定义参见GB 3096—2008。

8.6 抗时变能力

不同架构模式的移动设备声纹识别，当用户声纹时变导致在环境噪声等效声级≤40 dB(A)时，仍无法达到8.1中一般级的性能指标，则应要求用户重新进行声纹建模。

8.7 系统响应时间

移动设备声纹识别的系统响应时间应满足以下要求：

a) 进行声纹注册时，从移动设备提示用户已成功录制最后一条声纹语音样本起，到移动设备提示注册结果为止，所需时间小于或等于3 s；

b) 进行声纹比对时，从移动设备提示用户已成功录制声纹语音样本起，到移动设备提示识别结果为止，所需时间小于或等于2 s。

9 安全要求

9.1 基本要求

移动设备声纹识别的基本信息安全应符合GB/T 40660—2021中第4章的要求。

9.2 声纹数据采集安全

移动设备声纹数据采集的信息安全应符合GB/T 40660—2021中第5章的要求。

9.3 声纹数据存储安全

移动设备声纹数据存储的信息安全应符合GB/T 40660—2021中第6章的要求。

9.4 声纹数据使用安全

移动设备声纹数据使用的信息安全应符合GB/T 40660—2021中第7章的要求。

附 录 A
（资料性）
移动设备声纹识别典型应用模式

A.1 概述

本附录对移动设备上声纹识别技术架构的一些典型应用模式进行描述。在移动设备中，本附录的描述进一步细分为富执行环境和可信执行环境。

A.2 典型应用模式

A.2.1 结合可信执行环境的本地识别模式

图 A.1 描述了典型模式一。在这种模式下，声纹特征项提取、声纹建模、声纹数据存储和声纹比对都位于移动设备中，出于安全性考虑，声纹识别系统的各种功能应由移动设备中可信执行环境进行保护。语音采集元件或允许由富执行环境和可信执行环境共享访问，或仅允许可信执行环境访问。

移动应用一般位于富执行环境，通过可信执行环境提供的对外接口调用声纹识别系统，声纹识别系统调用位于移动设备中的语音采集元件开展对声纹语音样本的采集，在进行质量判断、呈现攻击检测后提取声纹特征项，根据目的不同进行声纹注册或者识别过程，并将执行结果反馈给移动应用。

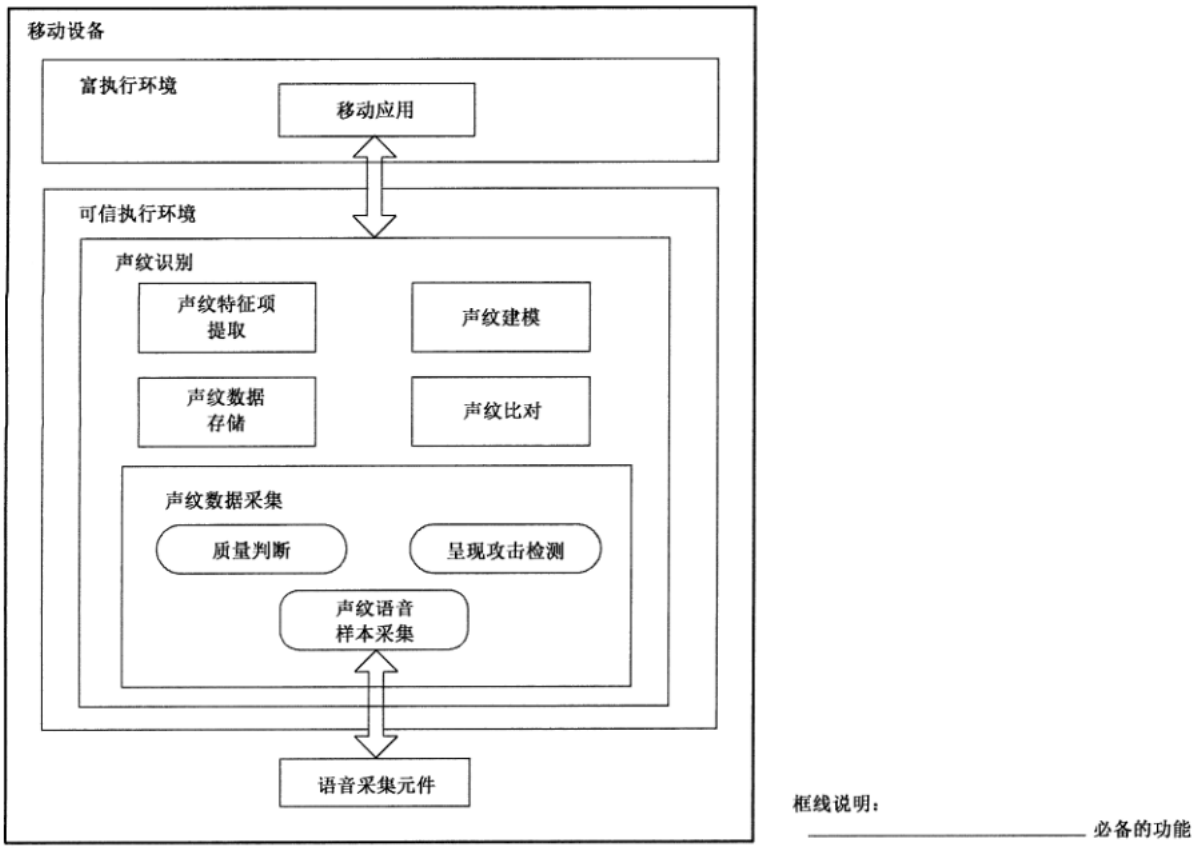

图 A.1 结合可信执行环境的本地识别模式

A.2.2 结合可信执行环境的远程识别模式

图 A.2 描述了典型模式二。这种模式下，声纹语音样本采集位于移动设备中，声纹建模、声纹数据存储和声纹比对在声纹服务器上完成，而声纹特征项提取则根据不同方案或位于移动设备端或位于声纹服务器端。出于安全性增强考虑，位于移动设备中的声纹语音样本采集和声纹特征项提取在可信执行环境中实现。语音采集元件或允许富执行环境和可信执行环境共享访问，或仅允许由可信执行环境访问。

移动应用一般位于可信执行环境，通过可信执行环境提供的对外接口调用声纹识别系统，声纹识别系统调用位于移动设备中的声纹语音样本采集功能开展对语音的采集，在进行质量判断、呈现攻击检测后进行声纹特征项提取，并访问声纹服务器进行声纹注册或识别过程，完成后向调用声纹识别的移动应用反馈结果。

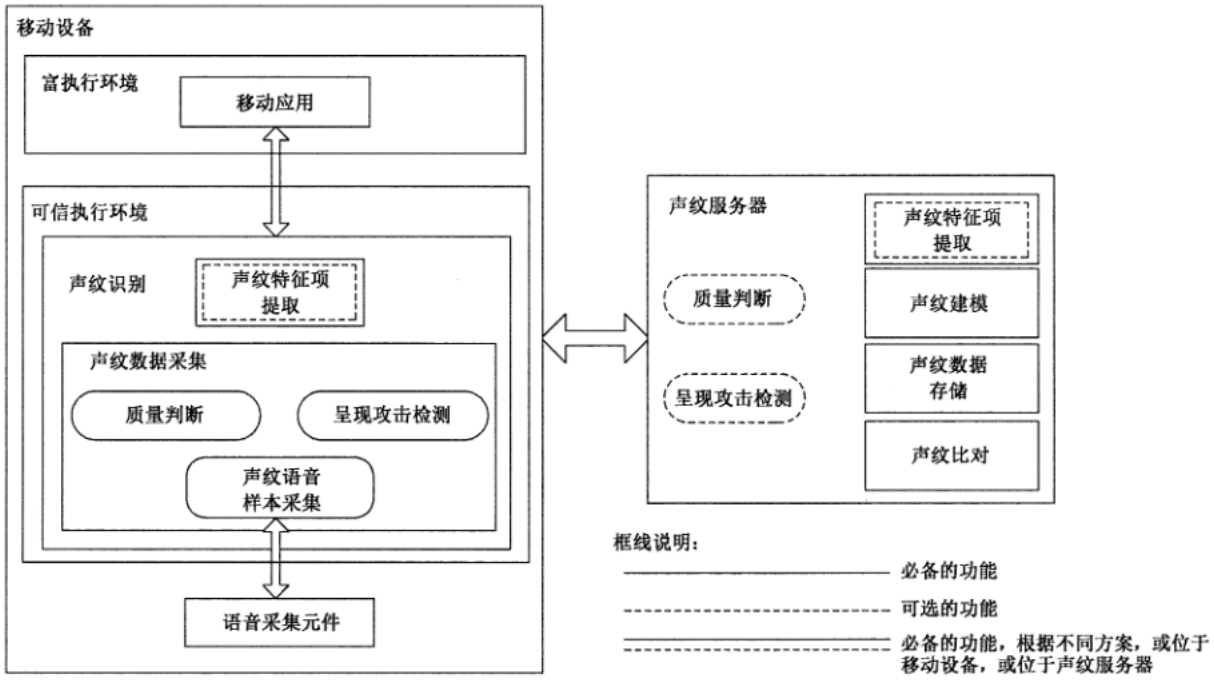

图 A.2 结合可信执行环境的远程识别模式

A.2.3 未结合可信执行环境的远程识别模式

图 A.3 描述了典型模式三。这种模式下，声纹语音样本采集位于移动设备中，声纹建模、声纹数据存储和声纹比对在声纹服务器上完成，而声纹特征项提取则根据不同方案或位于移动设备端或位于声纹服务器端。位于移动设备中的声纹语音样本采集在富执行环境中实现。语音采集元件允许通过富执行环境进行访问。

移动应用位于富执行环境，调用声纹识别系统后通过位于移动设备中的语音采集元件进行声纹语音样本的采集，在进行质量判断、呈现攻击检测后进行声纹特征项提取，并访问声纹服务器进行声纹注册或识别过程，完成后向调用声纹识别的移动应用反馈结果。

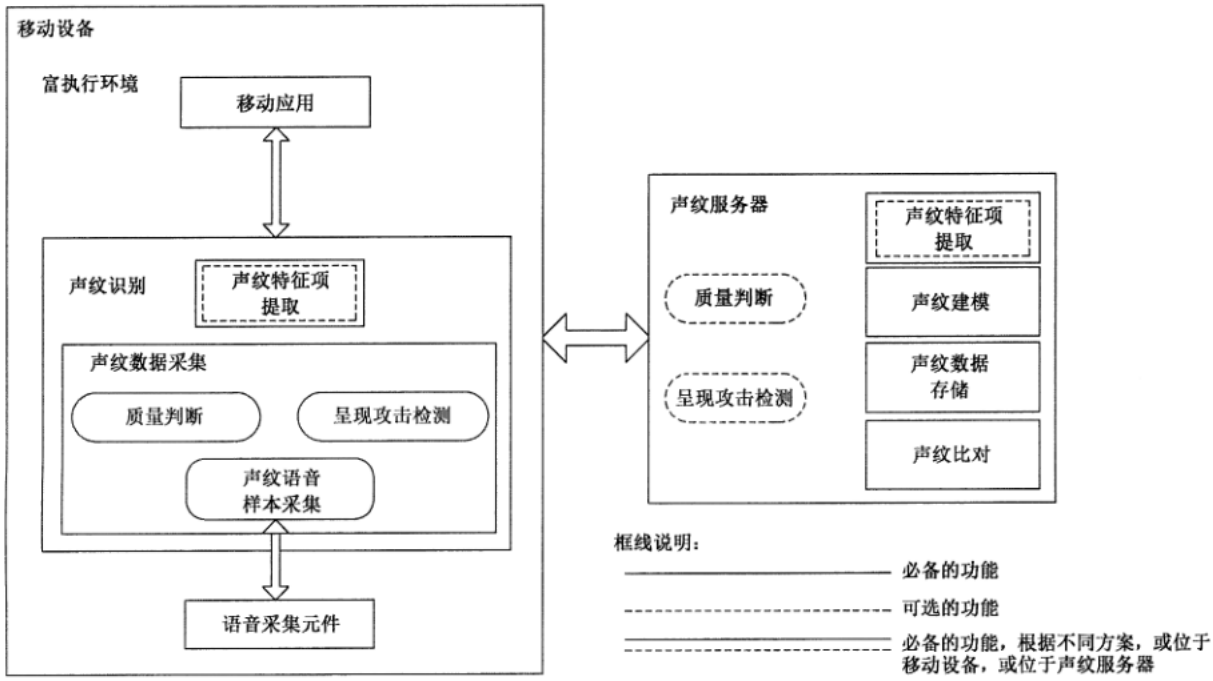

图 A.3　未结合可信执行环境的远程识别模式

附 录 B
（资料性）
移动设备声纹识别呈现攻击检测方法

移动设备声纹识别具备的呈现攻击检测功能可采用以下方法实现。

——针对远场偷录的重放语音的检测方法。由于远场偷录会造成频谱扁平化，且重放设备会造成频谱失真，因此提取频谱能量比值（Spectral Ratio，SR）、低频能量比值（Low Frequency Ratio，LFR）、调制系数（Modulation Index，MI）、子带调制系数（Sub-band Modulation Index，SMI）四种特征，并通过SVM进行建模，进行重放检测。

——使用常数Q倒谱系数（Constant-Q Cepstral Coefficients，CQCC）作为特征进行检测的方法。该特征在低频部分具有极高的频率分辨率，高频部分具有极高的时间分辨率，能更好地对重放失真进行表征。

——基于基频检测的方法。基于单元选择的方法做出来的合成语音，其基频会出现突变。基于统计参数估计的方法做出来的合成语音，其基频易过于平滑。因此可根据基频变化检测出合成方法，如基频变化过于平滑或出现突变，可判断是合成语音。

——基于相位信息的检测方法。人耳对语音的相位信息不敏感，在语音合成过程中相位信息会被忽略或不作针对性处理的。因此可根据语音的相位信息检测合成语音，如语音相位信息出现突变或数值范围过大，可判断为合成语音。

——基于相邻帧间动态变化性的检测方法。基于HMM的语音合成方法中，相邻帧之间的动态变化相对较小（更平滑）。因此可根据基频检测出合成方法，如帧间变化过于平滑，可判断是合成语音。

——基于频谱包络的检测方法。在一些合成语音中频谱包络趋于平滑，表现为高阶梅尔倒谱系数（Mel-frequency cepstral coefficients，MFCCs）方差较小。因此可根据高阶梅尔倒谱系数方差检测出合成方法，如高阶梅尔倒谱系数方差较小，可判是合成语音。

——基于频谱分辨率的检测方法。采用常数Q变换方法，低频用长窗，高频用短窗，对不同频率采用不同的分辨率。通过常数Q变换之后，语音在低频具有更高的频率分辨率，在高频具有更高的时间分辨率，对不同频率段进行观察可以找出合成语音与真实语音的差异。

参 考 文 献

[1] GB 3096—2008 声环境质量标准
[2] GB/T 5271.37—2021 信息技术 词汇 第37部分:生物特征识别
[3] GB/T 29268.1—2012 信息技术 生物特征识别性能测试和报告 第1部分:原则与框架