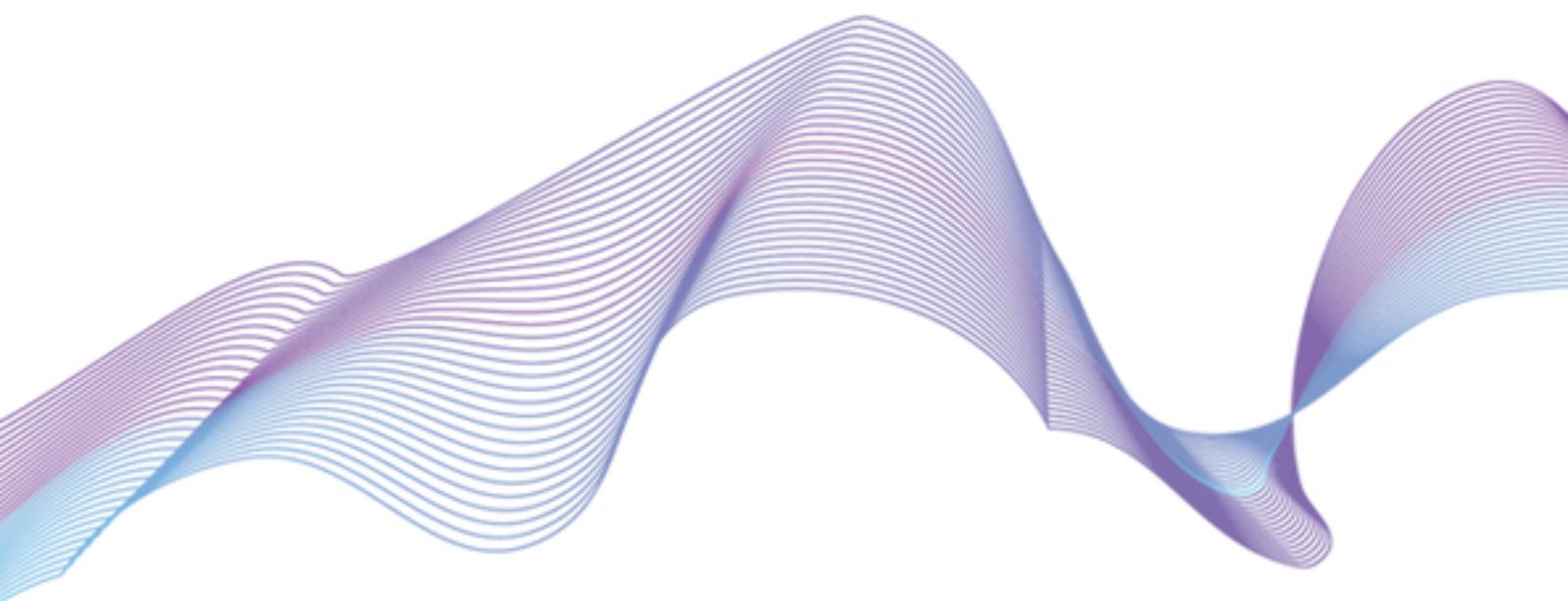


中国声纹识别产业发展白皮书 (2019)



联合发布

AIIA-得意音通人工智能声纹技术联合实验室
清华大学人工智能研究院听觉智能研究中心

2019年4月22日

编制声明

本白皮书版权属于人工智能产业发展联盟（AIIA）- 得意音通人工智能声纹技术联合实验室和清华大学人工智能研究院听觉智能研究中心。文中部分数据来源于网络公开资料整理，转载、摘编或利用其它方式使用本白皮书文字或观点的，应注明来源。违反上述声明者，编者将追究其相关法律责任。

主编

郑 方 孙明俊

编写组

成 舸 王 钰 李钰韩 程星亮 徐明星 王蕴韬

曹 峰 李 荪 宋丹峰 杨 悦 倪 鸣 陈柳村

周 鑫 邬晓钧 肖永明 向银杉 高 锐 刘彦全

序言

自上世纪四五十年代以来，经过长时间的探索实践，声纹识别技术终于逐渐走向成熟，迎来了历史上最佳的应用发展期。这一改变，首先发生在中国。

2018年可以说是声纹识别迈向产业化的分水岭，其标志性事件莫过于中国人民银行颁布的《移动金融基于声纹识别的安全应用技术规范》。这是我国金融行业第一个生物识别技术标准，其影响之深之广，远超我们的预期。它不仅为声纹识别技术进入金融领域突破了标准难题，也为多年来一直坚持初心的业界同仁注入了一针强心剂，2018年因此被称之为“声纹元年”。

声纹开始热了，产业态势已成，孤军奋战的日子一去不复返。再好的科研成果，只有从象牙塔里走出来，放到实践中去检验，才能发光发热，造福社会。作为行业老兵，我们感到欣慰，因为这证明了我们当初的判断和坚持是正确的。但我们同样也有担心，如果一个刚发芽的美好事物，被鱼龙混杂的信息误导、被不切实际的想象捧杀，对它的发展有害无益。

声纹识别是一项简单、优雅、安全的技术。我们希望这个朝气蓬勃的产业从一开始，就走上一条良性健康发展的道路，在开启新希望的同时，避免重蹈某些产业的覆辙。这就是我们为什么发布白皮书的初衷。

这份报告名为白皮书，实际上是从技术、产业、场景、企业、政策、资本等层面，根据可获得的一些公开资料，对声纹识别学术研究和产业发展脉络做了一份梳理，力求为广大研究人员、工程师、创业者、投资商和媒体人们提供一份快速了解行业概貌的“导览地图”。

我们力求尽量客观的反映这一新生行业的总体态势，但还谈不上多么全面和深入，里面提到的有些观点，也可能有不对或不妥之处，需要交给市场和时间去检验。

感谢 AIIA 对发布这份报告给予的支持。初试啼声，难免挂一漏万，还请业界朋友和广大读者多包涵，多指正。

是为序。

郑方

清华大学人工智能研究院听觉智能研究中心 主任

得意音通信息技术研究院 院长

目 录

| | |
|---------------------------------|-----------|
| 1. 声纹识别发展背景和环境分析 | 1 |
| 1.1. 移动万物互联，生物特征识别迅速抢占市场..... | 1 |
| 1.2. 人脸指纹漏洞频发，个人隐私泄露风险加剧..... | 2 |
| 1.3. 系列政策法规出台，国家加快 AI 产业引导..... | 3 |
| 1.4. 生物识别从生理特征走向行为特征..... | 4 |
| 1.5. 语音交互从千人一面向个性化进阶..... | 5 |
| 2. 声纹识别技术发展现状 | 6 |
| 2.1. 声纹是一种特殊的行为特征..... | 6 |
| 2.2. 声纹识别技术的分类逻辑..... | 7 |
| 2.3. 声纹识别技术进入商用..... | 9 |
| 2.4. 行业专利申请量激增..... | 11 |
| 3. 声纹识别产业发展现状 | 13 |
| 3.1. 市场规模：想象空间巨大，规模将超千亿..... | 13 |
| 3.2. 竞争格局：赛道选手增加，专业厂商领跑..... | 13 |
| 3.3. 市场份额：外企入华分羹，中企绝对占优..... | 13 |
| 3.4. 技术来源：技术门槛高企，来源高度集中..... | 14 |
| 3.5. 资本现状：融资进度提速，行业风口可期..... | 15 |
| 3.6. 公司案例：得意音通..... | 16 |
| 4. 主要应用领域及场景 | 17 |
| 4.1. 国家公共安全..... | 17 |
| 4.2. 泛金融..... | 18 |
| 4.3. 社会保险..... | 18 |
| 4.4. 智能安防及个性化语音交互..... | 19 |
| 4.5. 典型应用案例：金融领域..... | 19 |
| 4.5.1. 中国建设银行 | 19 |
| 4.5.2. 贵阳银行 | 24 |
| 4.5.3. 西安银行 | 25 |
| 4.5.4. 浦东发展银行 | 25 |

| | |
|---------------------------|-----------|
| 5. 标准化建设 | 27 |
| 5.1. 技术基础标准化..... | 27 |
| 5.2. 金融应用标准化..... | 29 |
| 5.3. 正在进行的标准化..... | 30 |
| 6. 技术挑战与发展趋势 | 32 |
| 6.1. 技术挑战..... | 32 |
| 6.2. 发展趋势..... | 35 |
| 附一：图表索引 | 37 |
| 附二：参考文献 | 38 |

1. 声纹识别发展背景和环境分析

1.1. 移动万物互联，生物特征识别迅速抢占市场

据 2019 年中国互联网络信息中心发布的《中国互联网络发展状况统计报告》显示，截至 2018 年 12 月，中国网民的规模达到了 8.29 亿，全年新增网民的数量是 5653 万，互联网的普及率是 59.6%，较前年底提升了 3.8 个百分点；中国手机网民的规模达到了 8.17 亿，全年新增手机网民的数量是 6433 万。截止去年 12 月，我国即时通信用户规模达 7.92 亿，网络新闻用户规模达 6.75 亿，网络购物用户规模达 6.10 亿，网上外卖用户规模达 4.06 亿，网络支付用户规模达 6.00 亿，网络视频用户规模达 6.12 亿，短视频用户规模达 6.48 亿。

以手机为中心的智能设备，成为“万物互联”的基础。移动支付、车联网、智能家居、智能安防等智能化应用场景多方位促进国人“衣食住行”体验升级。移动互联网服务场景不断丰富，应用场景日趋复杂，网络身份认证作为防护网络资产的第一道关口，在万物互联时代有着举足轻重的作用。

生物识别技术作为新一代人工智能的重要领域，借助人体生理特征或行为特征进行身份识别。近年来得益于云计算、大数据、物联网、深度学习等信息技术的快速发展，生物识别技术在基础理论、算法模型、创新应用、软件支持等方面不断取得突破。

根据前瞻产业研究院发布的《2018—2023 年中国生物识别技术行业市场调研与投资预测分析报告》显示，2013 年全球生物识别市场的规模达到 98 亿美元。2016 年生物识别技术行业的市场规模在 150 亿美元左右，到 2021 年这一数值将增加至 305 亿美元。2002 至 2015 年，国内生物识别市场的年复合增长率达到 50%，2016 年生物识别市场规模达到 120 亿元人民币左右。预计到 2021 年，中国生物识别行业的市场规模将突破 340 亿元人民币。



图表 1-1 2007—2021 年全球生物识别技术行业市场规模与预测

图表来源：前瞻产业研究院



图表 1-2 2002—2021 年中国生物识别技术行业市场规模与预测

图表来源：前瞻产业研究院

1.2. 人脸指纹漏洞频发，个人隐私泄露风险加剧

人脸识别安全漏洞在 2017 年的“3·15”晚会上首次曝光，凭借一张观众的自拍照成功“换脸”破解手机人脸识别系统。

➤ 假体攻击相关的安全漏洞：无效的活体检测

雷锋网：盗刷 28 万，某支付平台“人脸识别”现重大漏洞

<https://www.leiphone.com/news/201801/UfmGiSEhFiSVWBXo.html>

➤ 无意图检测带来的安全风险：“被指纹”和“被人脸”

凤凰网科技：最担心的事发生了！男子睡梦中被人用手机刷脸，醒来卡里一万多元没了！

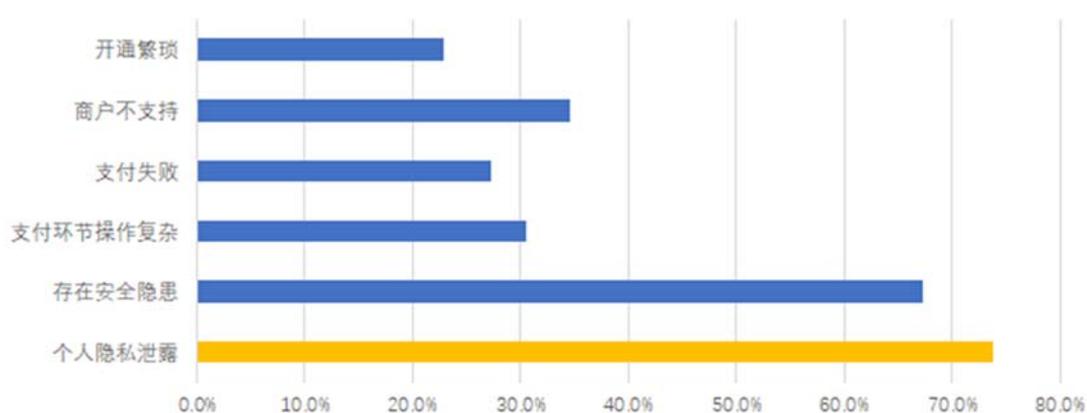
http://tech.ifeng.com/a/20190407/45588746_0.shtml

半岛都市报：女子网上交友被坑 熟睡中手机被男友盗刷 20 万

http://news.bandao.cn/news_html/201807/20180703/news_20180703_2842126.shtml

比起数据在传输和认证过程中的安全漏洞，后台的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。

中国支付清算协会 2018 年移动支付用户调研报告显示，“个人信息泄露是用户使用移动支付过程中最常遇到的安全问题”。在移动支付用户实际使用行为分析中，“个人信息被泄露”高居第一，占比高达 81.0%，高于第二位的“手机扫描到伪假条码” 70.1%



图表 1-3 移动支付用户对生物特征识别技术认知程度分析

数据来源：《2018 年移动支付用户调研报告》

1.3. 系列政策法规出台，国家加快 AI 产业引导

2017 年 6 月，《中华人民共和国网络安全法》正式施行，提出实施网络可信身份战略，支持研究开发安全、方便的电子身份认证技术，推动不同电子身份认证之间的互认；明确将个人生物识别信息纳入个人信息范畴进行规范管理。

2017 年 7 月，国务院印发《新一代人工智能发展规划》，从建立关键共性技术体系、探索行业创新应用等方面对生物识别技术发展提出了重点任务。

2017 年 12 月，工信部印发《促进新一代人工智能产业发展三年行动计划（2018-2020 年）》，智能身份识别系统、智能语音交互系统同时被列入率先取得突破的 8 大重点领域。

2018 年 9 月，第十三届全国人大常委会将《个人信息保护法》列入本届立

法规划。人工智能和大数据时代的个人信息保护即将迎来专门立法。

1.4. 生物识别从生理特征走向行为特征

生物特征识别最大的共性是唯一性。人的生理特征都存在唯一性，每个人都有独一无二的脸、指纹、虹膜等。由于每个人的生物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，所以利用生物识别技术进行身份认定相对于其他身份认证技术是安全且准确的。但也正是由于生理特征的不可撤销性，生物特征信息一旦被泄露、大量的带有唯一性的生物特征数据被盗取，基于生理特征的身份识别系统将彻底崩溃。这也是生理特征识别方式的真正‘痛点’。

| 远程身份认证技术 | 要素类别 | | | 金融领域场景示例 | 风险分析 | 安全等级 |
|---------------------------------------|------|------|------|--------------------------------|----------------------------|------|
| | 知悉要素 | 持有要素 | 生理要素 | | | |
| 手机号码+短信验证码 | √ | √ | | 手机银行找回密码 | 手机恶意软件窃取验证码 | 低 |
| 银行卡卡面信息+ 银行预留信息+ 交易密码+ 短信验证码 | √ | √ | | 非银行支付机构利用 银行卡绑定进行实名认证 | 银行卡卡面信息、交易密码、 手机短信验证码泄露 | 较高 |
| 身份证照片+人脸照片 | | √ | √ | 非银行支付机构利用 手持身份证照片进行 实名认证 | 照片被窃取或伪造 | 较高 |
| 指纹、人脸、虹膜等 生物认证技术 | | | √ | 手机银行登录、移动 支付 | 生物特征泄露、假体攻击 | 较高 |
| 身份证号+姓名+人脸 照片 | √ | | √ | Ⅱ、Ⅲ类账户远程开 户 | 身份证卡面信息泄露且人 脸照片被伪造 | 高 |
| 数字证书+PIN | √ | √ | | 网银专业版登陆 | 载体丢失且 PIN 泄露 | 高 |
| 声纹识别技术 (声纹+动态声纹密码) | | √ | √ | 手机银行登录、移动 支付 | 生物特征泄露、重放攻击 | 高 |

图表 1-4 移动金融中远程身份认证技术安全性比较

资料来源：中国人民银行《移动金融基于声纹识别的安全应用技术规范》标准编制组

1.5. 语音交互从千人一面向个性化进阶

2017 可以看做智能语音交互元年，各品牌厂商打造的智能音箱纷纷上市。2018 年语音交互落地突然加速，产品从智能音箱扩展到其他品类，如电视盒子、闹钟、灯、智能马桶等。未来，语音交互将迎来更多成长机会和更大的发展空间。在由全球最大中文 IT 社区 CSDN 组织编撰的《2018 中国人工智能产业路线图》中，将语音交互划分为三种阶段：

- L1 阶段：能以极高的准确率，在典型环境下响应用户的语音输入；
- L2 阶段：能以极高的准确率识别出交互的当事人和环境，然后进行个性化的交互；
- L3 阶段：只要有数据，语音交互系统的能力就可以无边界扩展（包括个性和能力）。

在 L2 阶段体现的是个性化，语音交互不再停留在千人一面的固定化模式中，而是实现千人千面的个性化交互体验。

注：以上段落援引自声智科技李智勇《语音交互的进阶之路：层级、技术与颠覆性》

2. 声纹识别技术发展现状

2.1. 声纹是一种特殊的行为特征

声纹，是对语音中所蕴含的、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，而声纹识别是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。与指纹类似，每个人在说话过程中所蕴含的语音特征和发音习惯几乎是独一无二的，即使是模仿，也难以改变说话者最本质的发音特性和声道特征。由于声音的特殊性，声纹识别与其他行为特征相比，又兼具生理特性。

声纹识别和语音识别在原理上一样，都是通过对采集到的语音信号进行分析和处理，提取相应的特征或建立相应的模型，然后据此做出判断。但二者的根本目的，提取的特征、建立的模型是不一样的。声纹识别试图寻找的是区别每个人的个性特征，而语音识别则是侧重于对话者所表述的内容进行识别。

简而言之，语音识别（Speech Recognition）关心说的什么（What），声纹识别（Voiceprint Recognition）关心谁说的（Who），声纹识别通常又称作说话人识别（Speaker Recognition）。

声纹特征以声音为载体，具有以下特点：

1) **交互性**。声音是唯一可双向传递信号的生物特征，既可以接收信息，也可以发出信息，实现交互。

2) **便捷性**。声音是唯一周边无死角的生物特征，可以实现非接触式采集，方便使用。

3) **变化性**。声音是高可变性与唯一性的完美统一。没有两个声音是完全一样的，但里面所蕴含的信息，比如你是谁、你的年龄、你的情感等信息却都是唯一确定的。这种高可变性和唯一性的完美统一使得语音信号自身就具备了很强的防攻击能力。

4) **丰富性**。声音有“形简意丰”的特点，它虽然只是一个一维信号，但是蕴含着丰富的信息。在相同一段语音中，除了包含说话人信息外，还包含内容、语种、性别、情绪、年龄，甚至包含出生地、身体健康状况等丰富的信息。

声纹识别作为生物特征识别技术中的一种，作为一种身份认证手段，具有如

下优势：

1) **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）”。

2) **难以伪造**。声纹蕴含于人的语音当中，而语音是千变万化的。即便人两次读相同的内容，也不可能发出完全相同的声音。正是声纹这种“蕴不变于千变万化之中”的特性，使得声纹特征更加深层、难以琢磨，伪造起来也更为困难，因而认证强度更高、更安全。

3) **隐私性弱**。俗话说“身体发肤受之父母”，在一般的大众认知中，人们在很多场合下往往不愿意被拍照或者按手印，但采集一段随机跟读的声音，对于大众更加容易接受。

| | 错误接受率 | 错误拒绝率 /% | 容易实用性 | 处理速度人 | 易仿冒程度 | 生物特征 |
|-----|-------|-------------|----------------------|-------------------------------|-------|------|
| 指纹 | 很低 | 较低 | 好 | ≤1s | 一般 | 生理特征 |
| 掌纹 | 低 | 5 | 使用困难 | 5-10s | 一般 | 生理特征 |
| 虹膜 | 很低 | 约 10 | 需要培训才能使用, 手工操作对虹膜有困难 | 一起自动对准虹膜需要 3-5s, 手工操作需要 5-25s | 极难 | 生理特征 |
| 视网膜 | 未知 | 未知 | 不好 | 15-30s | 极难 | 生理特征 |
| 人脸 | 低 | <0.2 | 非常好 | ≤5s | 一般 | 生理特征 |
| 声纹 | 低 | 低 | 可以 | 1-3s | 难 | 行为特征 |
| 签名 | 低 | 10 | 一般 | 5-10s | 一般 | 行为特征 |

图表 2-1 生物特征识别技术特性对比

资料来源：《信息安全研究》2016 年 1 月第一期

2.2. 声纹识别技术的分类逻辑

声纹识别技术是一个统称，实际上，按照不同的应用方式和其他的一些限定，声纹识别又被分成了多个技术类别。其中，按照实际应用的方式可分为四类：

1) 声纹确认

即给定一个说话人的声纹模型和一段只含一名说话人的语音，判断该段语音

是否是该说话人所说。

2) 声纹辨认

即给定一组候选说话人的声纹模型和一段语音，判断该段语音是哪个说话人所说。

3) 声纹检出

即给定一个说话人的声纹模型和一些语音，判断目标说话人是否在给定的语音中出现。

4) 声纹追踪

即给定一个说话人的声纹模型和一些语音，判断目标说话人是否在给定的语音中出现，若出现，则标示出对话语音中目标说话人所说的语音段的位置。

正因为声纹是一种动态的行为特征，在上述普遍适用于各种生物特征识别技术的分类之外，声纹识别还多了一个技术维度——它的分类与说话的内容有关。

（注：还有一种分类是语言相关性：语言无关、语言相关。详见《自动声纹识别（说话人识别）技术规范》（2008年3月工信部（原信息产业部）发布）

根据声纹识别与待识别语音的文本内容的关系，声纹识别又可分为三类：

1) 文本无关

即对于语音文本内容无任何要求，说话人的发音内容不会被预先限定，说话人只需要随意录制达到一定长度的语音即可。这种方法使用起来更加方便灵活，具有更好的推广性和适应性。

2) 文本相关

即要求用户必须按照事先指定的文本内容进行发音。由于文本相关场景下，语音内容受到限定，整体随机性比文本无关场景下的小，所以一般来说其系统性能也会相对好很多。

3) 文本提示

即从说话人的训练文本库中，随机提取若干词汇组合后提示用户发音。既对语音内容的发音范围进行了限定，又通过随机组合的方式，保留了语音内容的随机性，是文本无关与文本相关的一种结合。这种方式能一定程度上避免文本相关时的假冒录音闯入问题，同时具有较高的系统性能，且实现方便，是说话人识别

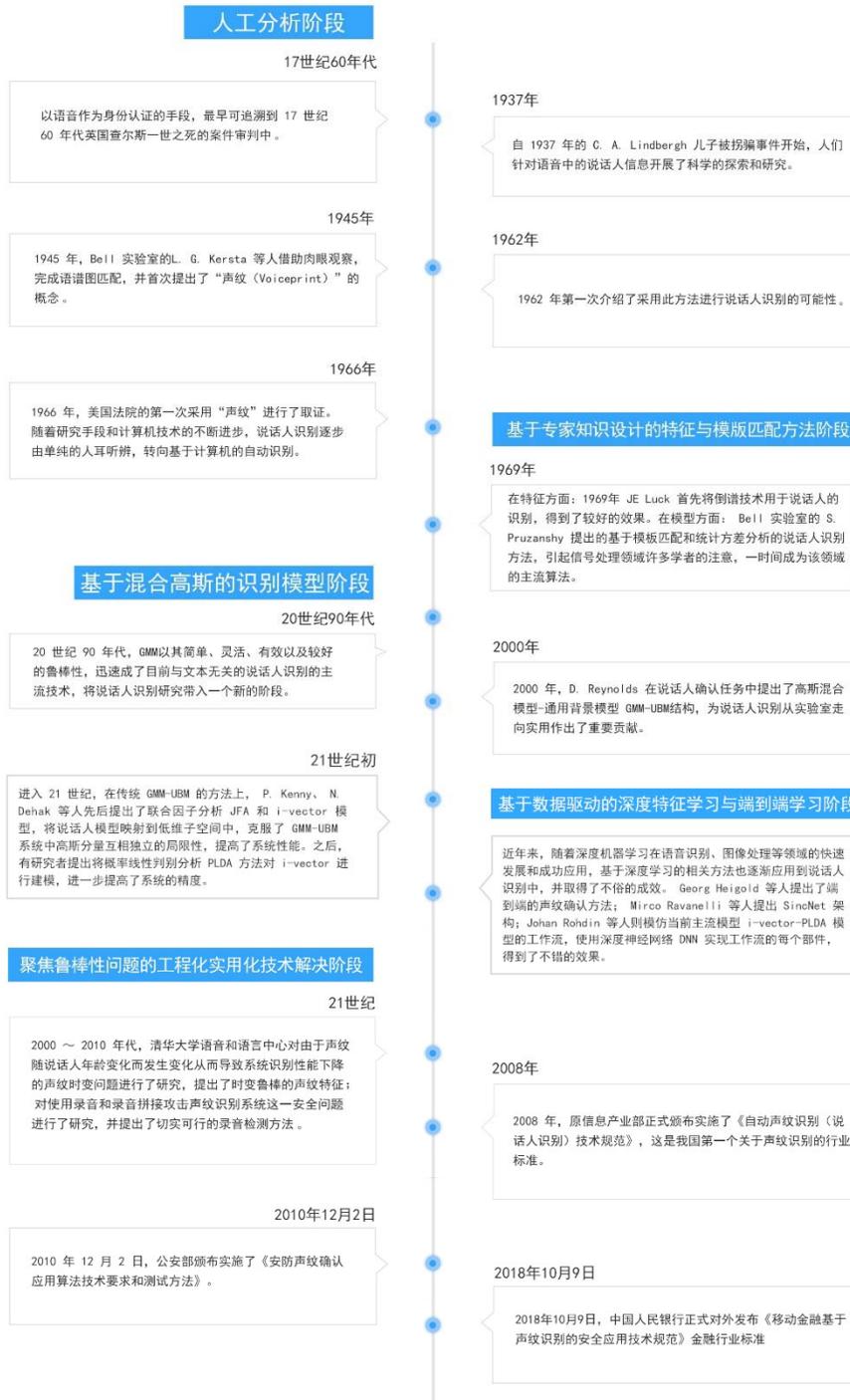
技术的一大热点。

不同类别的声纹识别技术之间，其采用的算法也会有细微的不同，对应的应用领域也会有所不同。无论是哪种声纹识别技术，都经过了漫长的发展过程。

2.3. 声纹识别技术进入商用

近年来，依托算法迭代、模型升级等优化措施，在识别说话人的共振峰、基音、倒频谱等声学特性方面取得一定突破，初步解决识别准确率、稳定性等问题。目前，声纹识别技术已支持对千万级以上容量的声纹库开展秒级检索识别。

声纹识别技术发展史

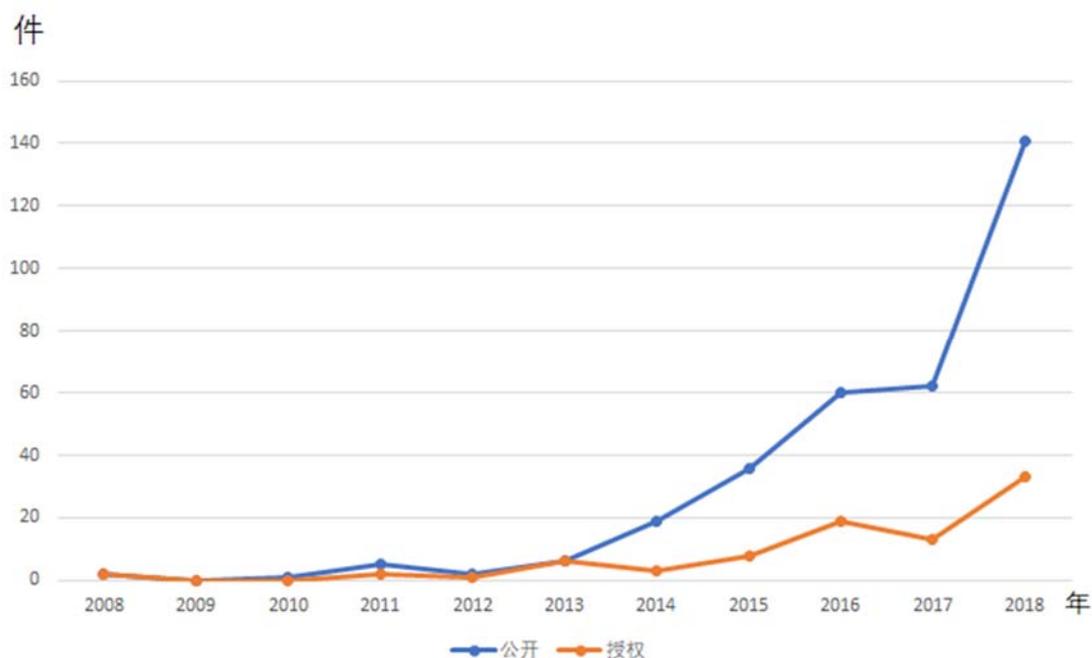


图表 2-2 声纹识别技术发展简史

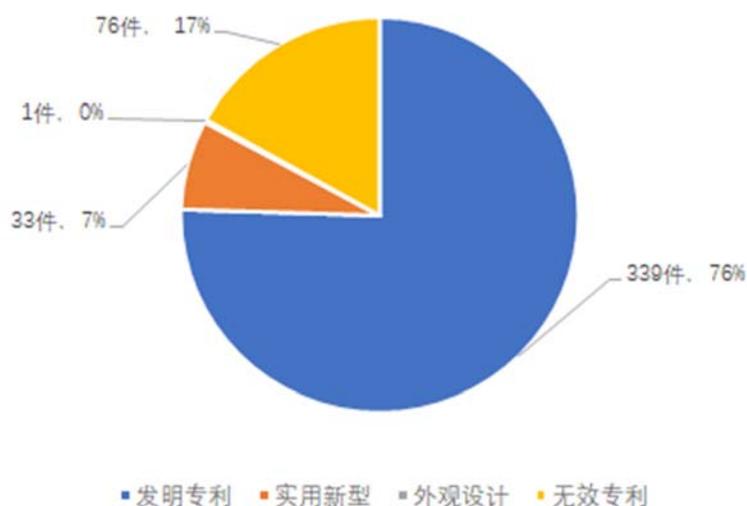
资料来源：《声纹识别：走出实验室 迈向产业化》

2.4. 行业专利申请量激增

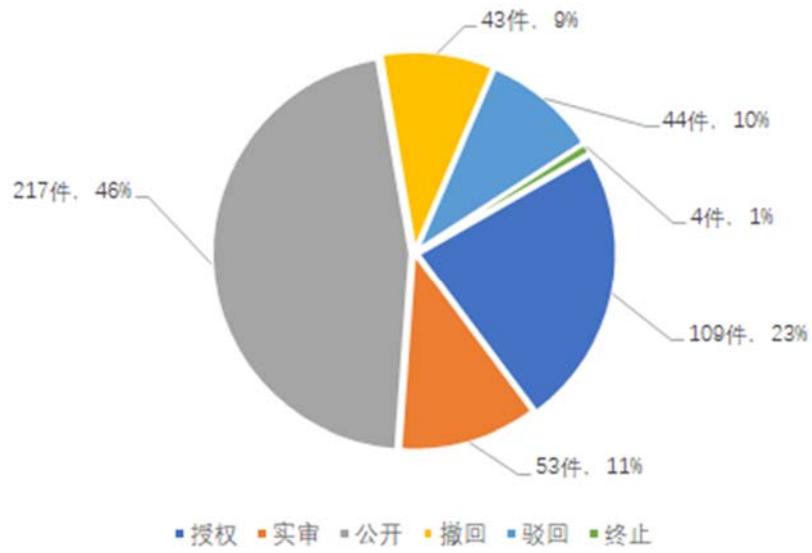
2013年起，国内相关专利公开数量呈大幅上升，5年内翻了10倍以上。相对专利公开数量，专利授权数量相对增长较缓，总数不超过40件。2018年，无论公开数量还是授权数量，专利增幅均达历史峰值。



图表 2-3 国内声纹相关专利申请数量宏观态势



图表 2-4 国内声纹相关专利类型分布



图表 2-5 国内声纹相关专利法律状态

数据来源：国家知识产权局数据库

3. 声纹识别产业发展现状

3.1. 市场规模：想象空间巨大，规模将超千亿

据前瞻产业研究院的分析，当下全球生物识别产业规模庞大，仅语音生物识别（注：即声纹识别）这一细分方向的市场规模就将近百亿美元，预计 2020 年更是有望超过 200 亿美元(合 1346 亿元人民币)，占整个生物识别市场的 22.4%。

从网络身份认证应用领域来看，据国外权威调研机构 MarketsandMarkets 数据显示，2019 年网络安全市场预计增长至 1557.4 亿美元，其中，身份认证信息安全市场规模将超过 300 亿美元。声纹识别在其中也将扮演重要角色。

3.2. 竞争格局：赛道选手增加，专业厂商领跑



图表 3-1 中国声纹识别相关厂商成立时间轴

数据来源：相关企业官网

从时间轴上可以看出，可大致分为三个阶段：

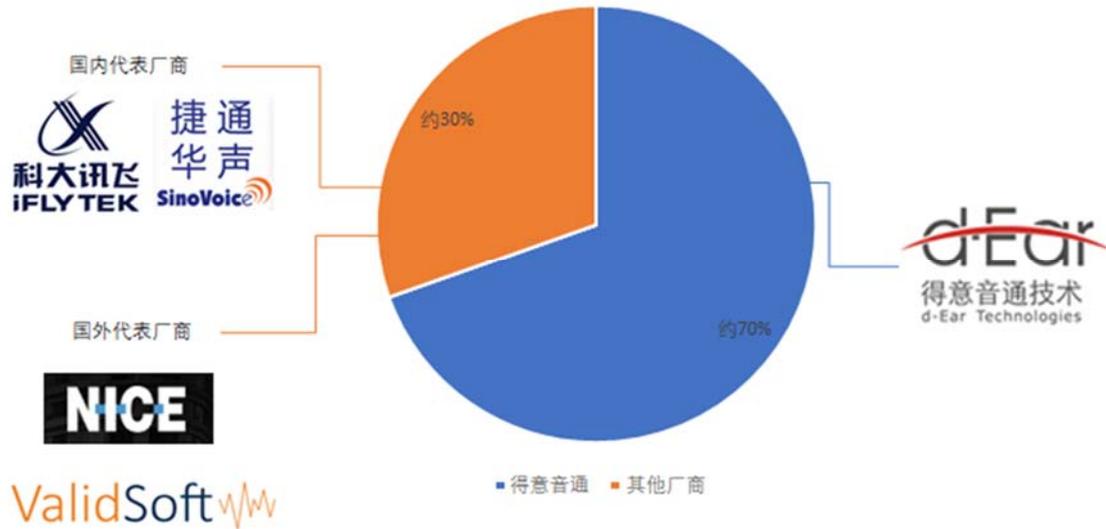
2000 年前后：第一批公司成立，主要是老牌语音厂商。2005 年前后：第二批公司成立，主要是深耕声纹的专业厂商。2015 年前后：第三批公司成立，主要是贴着 AI 标签的创业公司。最近两年，BAT 等大公司也相继在声纹领域开始布局。

3.3. 市场份额：外企入华分羹，中企绝对占优

2018 年 5 月，花旗银行宣布正式在中国市场推出“声纹验证”服务，其技

术供应商为曾为苹果提供 Siri 技术支持的国际语音行业老牌厂商 Nuance。

之后，ValidSoft、Nice 等外企开始进入中国金融业市场，但市场份额有限，总体来看，国内金融市场的声纹供应商里，以得意音通为代表的中资企业占有绝对优势。

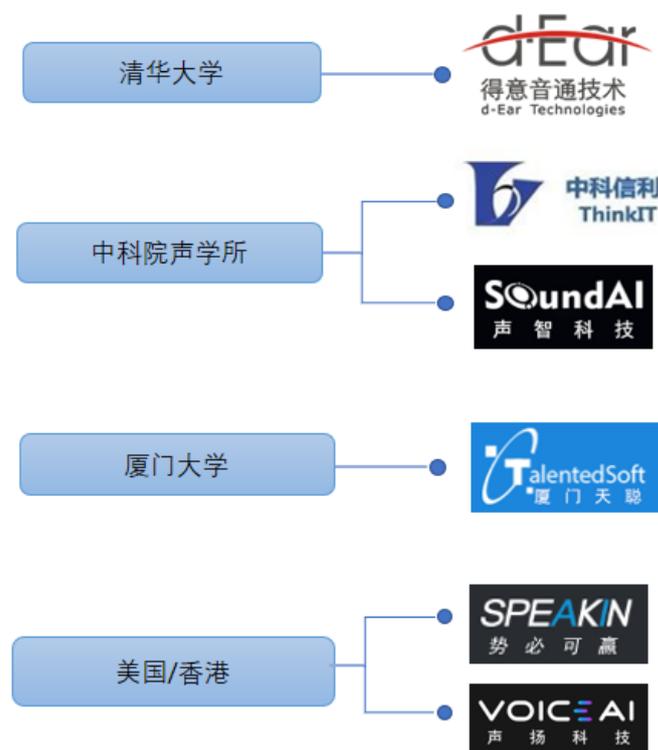


图表 3-2 中国移动金融市场上的主要声纹识别厂商分布

3.4. 技术来源：技术门槛高企，来源高度集中

最近两年来，随着市场逐渐兴起，声纹识别相关厂商迅速从个位数扩充到二位数，其中既有深耕多年的专业声纹或语音厂商，也开始涌现出一批新面孔。

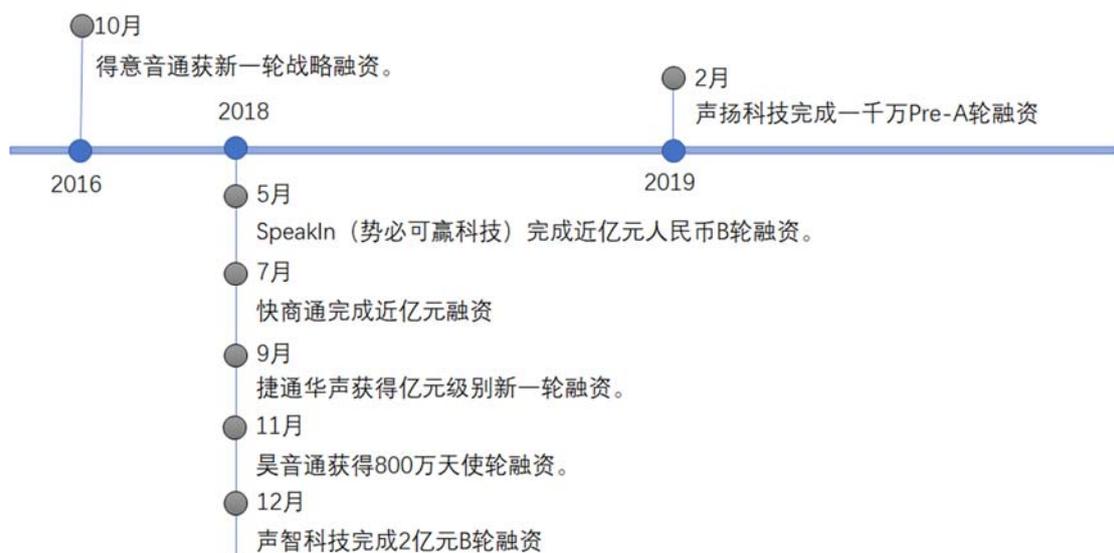
不过，在专业声纹厂商中，有明确技术来源的厂商总体占比不高，主要集中在少数几家有深厚积淀的高校和科研机构；相当一部分存在于搜索引擎和媒体上的厂商，技术来源不够清晰。



图表 3-3 各声纹技术相关厂商技术来源

数据来源：相关企业官网

3.5. 资本现状：融资进度提速，行业风口可期



图表 3-4 中国声纹识别相关企业最新融资状况

3.6. 公司案例：得意音通

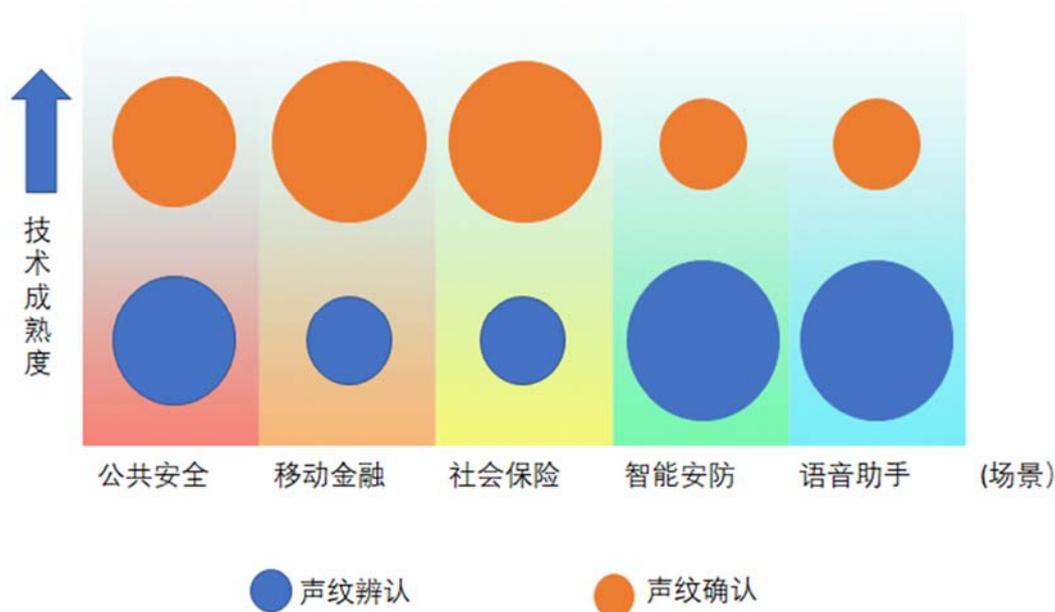
北京得意音通技术有限责任公司（简称“得意音通”），2002 年成立，是清华大学知识产权入股的高科技企业、中国人工智能产业发展联盟（AIIA）理事单位、互联网金融身份认证联盟（IFAA）理事单位，是拥有完全自主知识产权的声纹识别（VPR）、语音识别（ASR）与中文自然语言处理（CNLP）领域的技术研发商、产品开发商和服务提供商。得意音通在声纹识别这一被美国列为战略安全技术的领域居国际领先地位，可提供高安全、弱隐私、低成本的无监督身份认证服务。

得意音通核心技术源自清华，2015 年与清华大学成立“清华大学-得意音通声纹处理联合实验室”，2017 年组建得意音通信息技术研究院，延聘全球顶级人工智能专家，由张钹院士（中）和庄炳焯院士（美）任联席主任。其“基于动态密码语音的无监督身份认证系统”于 2018 年 5 月被中国电子学会组织的科技成果鉴定评价为“整体处于国际领先水平”。在国际音频情感识别竞赛 MEC 2017 中获冠军；在国际自动说话人验证欺骗和对策挑战赛 ASVspoof 2019 中获冠军。

得意音通核心技术完全自主可控，拥有十六项国家发明专利和一项国际发明专利。“声密保”内核融合声纹识别和语音识别，集成防录音攻击、防时变、情感识别等进行无监督身份认证，其产品和服务已首先成功应用于中国建设银行手机银行，并以“零事故”和“零投诉”通过数亿次验证。目前，已在国家信息中心、中国银联、中国互金协会、贵州省大数据中心、包括中国建设银行在内的多家银行、包括贵州社保在内的多省社保等成功应用，在各行业布局初成。

得意音通主导起草了我国第一个声纹识别标准，以及截至目前声纹识别领域所有的国家和行业标准。尤其是主导了金融领域声纹识别安全应用技术标准的起草，该标准已于 2018 年 10 月 9 日由中国人民银行颁布至各银行、证券、保险、基金，以及非银行支付机构，它也是我国金融领域第一个生物特征识别安全应用技术标准，得意音通拥有标准中规定的动态密码语音的相关发明专利。

4. 主要应用领域及场景



图表 4-1 声纹识别技术分支-应用场景分布图（圆圈越大代表成熟度越高）

4.1. 国家公共安全

随着固定电话和移动通讯网络的发展，声纹辨认技术首先在针对特定人群的国防安全、公安技侦、司法矫正等领域投入使用，有力保障了国家和公共安全。

例如在战场环境下，声纹辨认技术可察觉电话交谈过程中是否有关键说话人出现，一旦通过电话发出军事指令时，便可对发出命令者进行身份辨认（敌我指战员鉴别）。据报道，2001年4月1日迫降在我国海南机场的美军EP-3侦察机就载有类似的声纹识别侦听模块。在反恐作战中，恐怖分子在作案前后的通讯中往往会包含关键内容，因此，在通信系统或安全监测系统中预先安装声纹辨认系统，可通过通讯跟踪和声纹辨别技术对罪犯进行预防和侦查追捕。据悉，拉登的落网正是美国情报部门充分利用了声纹鉴别技术。

此外，声纹辨认技术还用于对满刑释放的犯罪嫌疑人进行监听和跟踪，可有效阻止犯罪嫌疑人再次犯科；针对通过电话勒索、绑架等刑事犯罪案件，公安司法人员也可利用声纹辨认技术，从通话语音中锁定嫌疑犯人、缩小刑侦范围。

4.2. 泛金融

随着网上支付、手机支付等成为现代人购物付款的主流方式，网络支付的身份认证开始愈发重要。为防止盗刷等案件发生，将声纹确认技术加入到交易支付中，通过动态声纹密码的方式进行客户端身份认证，可有效提高个人资金和交易支付的安全。

在国外，英国巴克莱银行、美国花旗银行、澳大利亚国家银行、万事达卡机构等都已开始引入声纹技术。在我国，2016到2018年间，中国建设银行、贵阳银行、兰州银行、西安银行、中国银联、中国互金协会等多家单位上线了声纹识别身份认证服务，用于账户登录、大额转账、无卡取款、密码找回等业务场景。微信和支付宝也已上线基于声纹动态口令的登录方式。此外，在信贷业务中引入声纹识别技术作为反欺诈手段，还可有效降低冒用他人身份进行骗贷以及多头贷款等事件的发生率。

4.3. 社会保险

城乡养老保险是社会保障体系的重要组成部分，目前我国社保面临三个问题：加速进入老龄化社会，社保金管理存在冒领问题，现有解决冒领问题的方法成本高、效率低。利用声纹识别技术有效解决参保人员的远程和现场的身份认证问题，避免了指纹和人脸等需现场办理、不易采集、易伪造等问题，杜绝了身份造假的可能性，节省大量成本，降低养老金冒领有效杜绝了养老金流失，同时方便老年人足不出户安全快捷领取养老金。以往半年甚至一年一次的认证，可以提高为每月一次；以往的参保用户抽查方式，可以变为全面普查，为社保机构免去诸多的的人力成本、物力成本、行政成本和时间成本。

声纹生存认证系统目前正在全国多省及自治区进行试点工作，为多地群众百姓提供了极大的便利。



图表 4-2 中国声纹社保试点分布图

4.4. 智能安防及个性化语音交互

随着语音技术的普及，越来越多的声纹识别应用场景还在不断涌现。如利用声纹确认技术，可完成个人日常生活中的各种事物访问控制的授权，比如智能手机锁屏、各类网络账号的声控密码锁、电脑声控锁、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。

目前这些新兴需求大部分还处在探索阶段。

4.5. 典型应用案例：金融领域

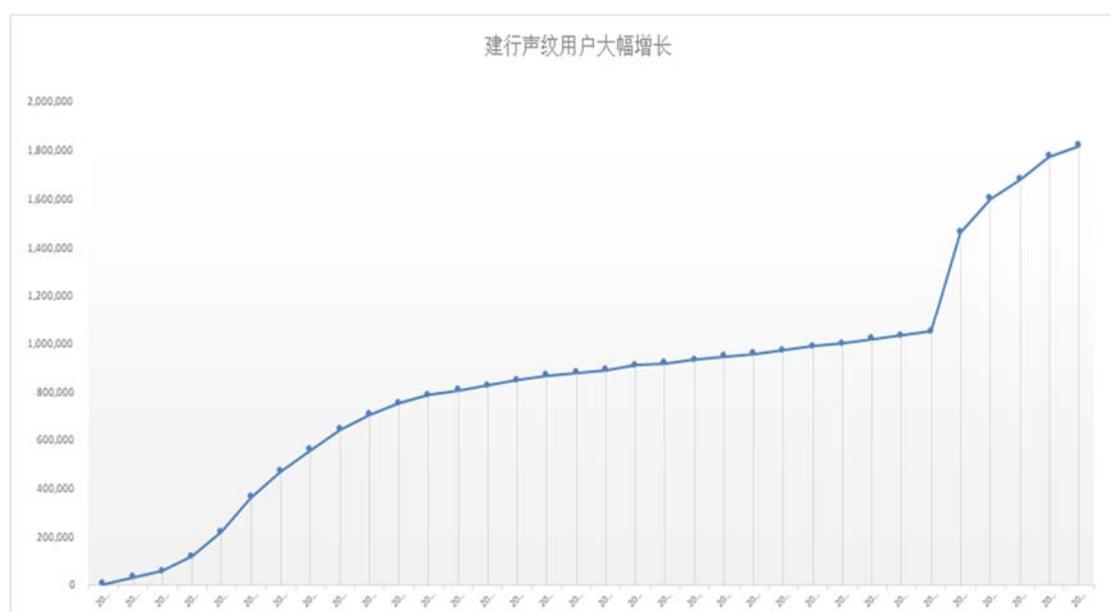
4.5.1. 中国建设银行

2016年5月，中国建设银行与清华大学合作率先推出手机银行中的声纹识

别应用，并在建设银行手机银行 APP 中正式上线。作为传统认证方式的一项增强安全手段，被应用于建行手机银行 APP 登录、转账、取款、支付等几乎所有环节。该系统上线以来，经由用户口口相传，目前在线有效用户数已超过 100 万，调用声纹识别的业务笔数逾 2 亿次，并创造了“零差错”和“零投诉”的记录。

1) 发展规模

据中国建设银行官方最新数据显示，中国建设银行声纹注册用户模型数增长到 182 万个，总交易量达到 2.4 亿次。



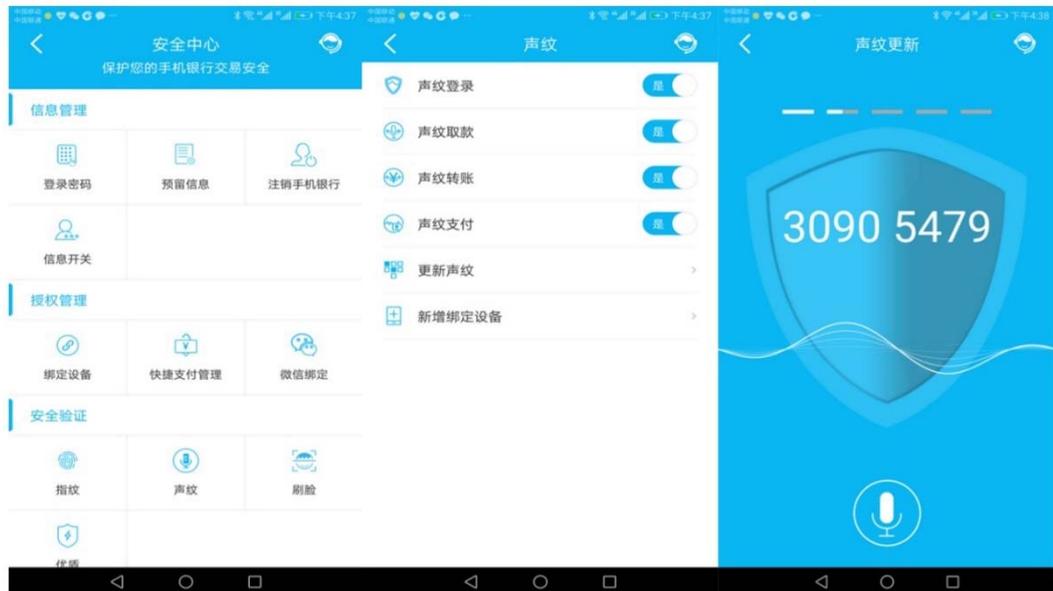
图表 4-3 中国建设银行声纹注册用户模型数增长变化

2) 功能开通及设置

用户登录手机银行 APP，可以在个人安全中心，进行安全认证设置，可以选择各个应用场景时，需要进行的身份识别技术，比如登录场景要求用户选择进行双因子认证，就是从多种身份识别技术中选择两种进行双重认证，从而提高了账户的安全级别。

如图所示，登录手机银行 APP，进入安全更新，选择声纹，第一次使用需要用户进行声纹预留，选择更新声纹，即进入声纹预留界面。

进行声纹建模应选择按键的场景，和手机扬声器的距离保持在 20cm 左右，声音平缓的念出显示的 8 位随机数字，一共进行 5 遍，即完成用户的声纹建模。

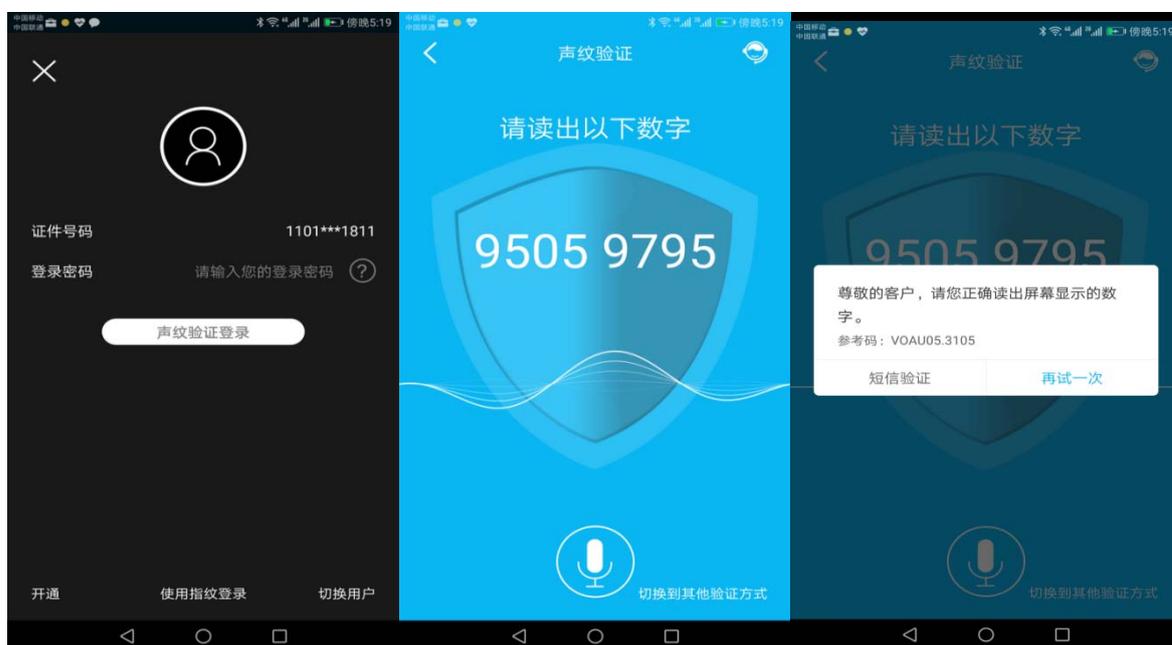


图表 4-4 建行手机银行 APP 声纹建模/更新界面

3) 应用场景

3-a) 手机银行登录

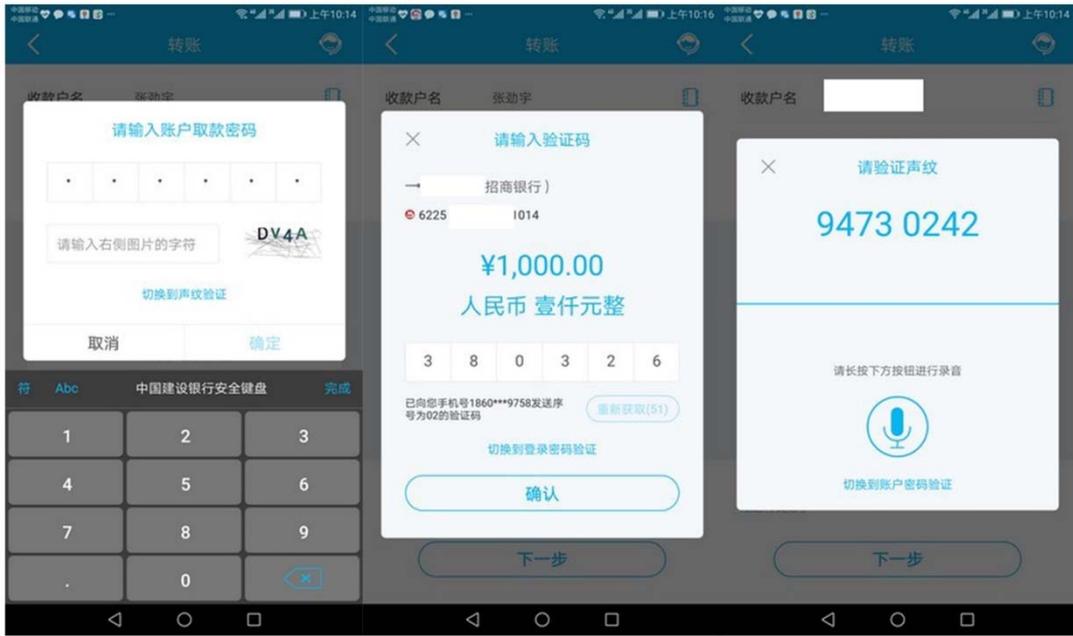
以建设银行手机银行 APP 为例，在完成开通“声纹验证”功能并预留用户个人声纹信息后，当用户再登录手机银行时，就会要求进行多重验证，进行用户名密码登录或指纹识别登录，通过后进入到“声纹验证”页面，如果进行声纹验证时读错显示的 8 位数字，或现场环境过于嘈杂，无法顺利使用声纹验证，此时可以选择重读（随机数字会重新生成）或切换到短信密码认证，通过了双重验证通过后可登录手机银行 APP 进行操作（见下图）。



图表 4-5 建行手机银行 APP 用户双重身份认证登录

3-b) 手机银行转账/支付

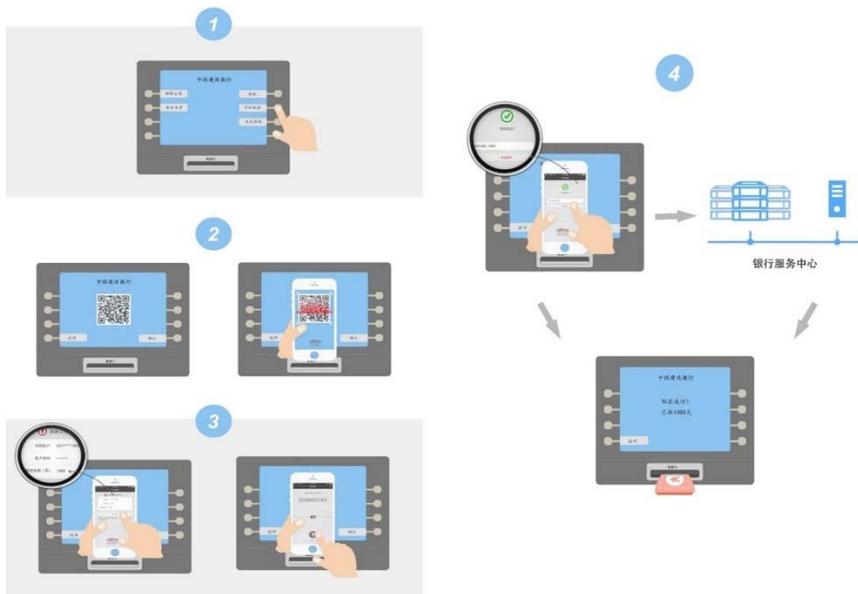
除了在手机银行登录时使用声纹识别进行双重身份认证登录时，开通声纹验证功能的账户在通过手机银行转账时，在转账金额达到一定数额或第一次给某账户转账时，除了默认的短信验证码之外，可以选择进行声纹验证以加强对用户账户资金的保护。



图表 4-6 建行手机银行转账声纹验证过程

3-c) ATM 机无卡取款

中国建设银行推出的声纹取款功能，利用最新的声纹识别技术，通过将客户声纹与其在系统中预留的声纹进行比对，来代替银行卡和密码验证，实现无卡轻松取款。

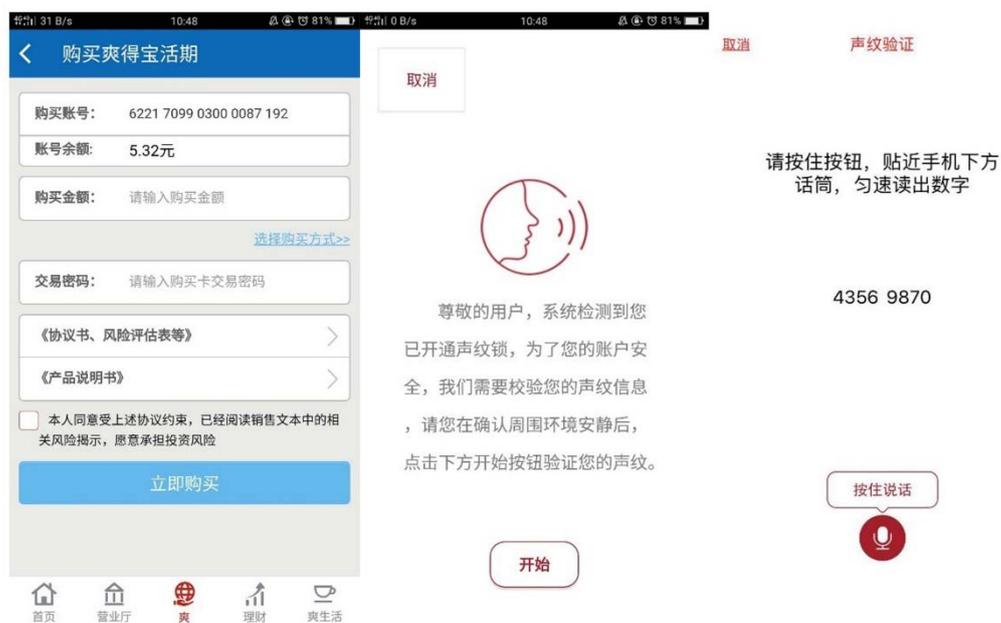


图表 4-7 建行手机银行声纹取款业务流程示意图

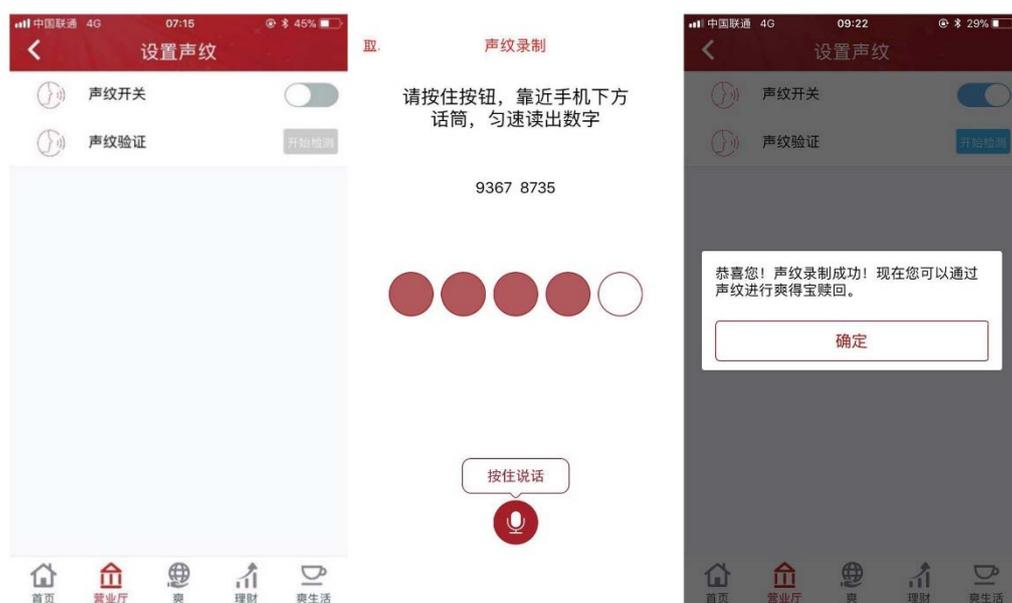
4.5.2. 贵阳银行

2016年8月15日，得意声密保产品与贵阳银行的业务系统完成联调并正式上线，应用于贵阳银行手机银行用户的登录与交易。

应用场景：通过移动应用购买理财产品时的确认，规划应用到登录、转账场景。



图表 4-8 贵阳银行手机银行理财产品交易声纹验证



图表 4-9 贵阳银行手机银行声纹注册界面

4.5.3. 西安银行

2018年11月23日，得意声密保软件在西安银行手机银行系统中正式上线投产，为西安银行手机银行用户提供登录、交易等场景的声纹识别身份认证服务。声纹识别主要应用在西安银行手机银行的登录、支付、转账等场景中，全面的保障用户资金安全，手机银行中使用的是声密保3.0系统，具备防录音、防时变、抗噪音等性能，各项性能指标均符合央行标准规范。这也是在央行发布《移动金融基于声纹识别的安全应用技术规范》后，第一个符合声纹标准的声纹识别技术应用。

应用场景：登录身份核实，规划应用到转账理财赎回等各种核身场景。

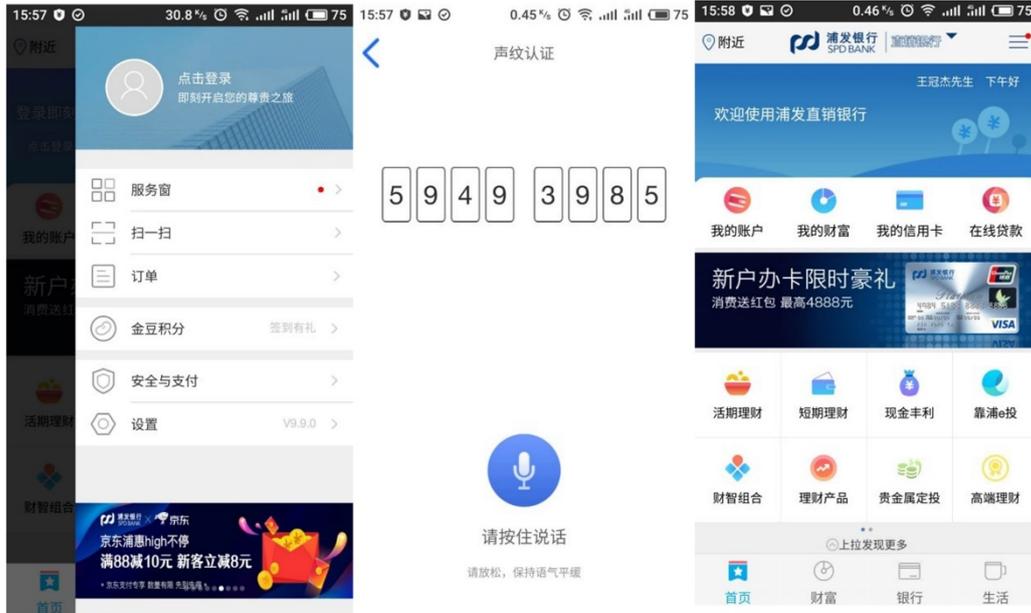


图表 4-10 西安银行手机银行声纹登录场景

4.5.4. 浦东发展银行

2018年3月底完成了“得意声密保软件”在生产环境的部署，开始白名单试用。2019年2月底，完成了全生产线部署工作实现生产线上的全面正式应用。

应用场景：手机银行免密登录，规划应用在转账、理财、支付等各种业务交易的身份认证环节。



图表 4-11 浦发银行手机银行声纹登录



图表 4-12 浦发银行手机银行声纹注册界面

5. 标准化建设

5.1. 技术基础标准化

2008年3月，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》。这是中国颁布的第一个关于声纹识别的标准。



图表 5-1 《自动声纹识别（说话人识别）技术规范》

2014年8月，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》。这是中国首次就声纹识别的应用技术要求和安全等级制订标准。



图表 5-2 《安防声纹确认应用算法技术要求和测试方法》

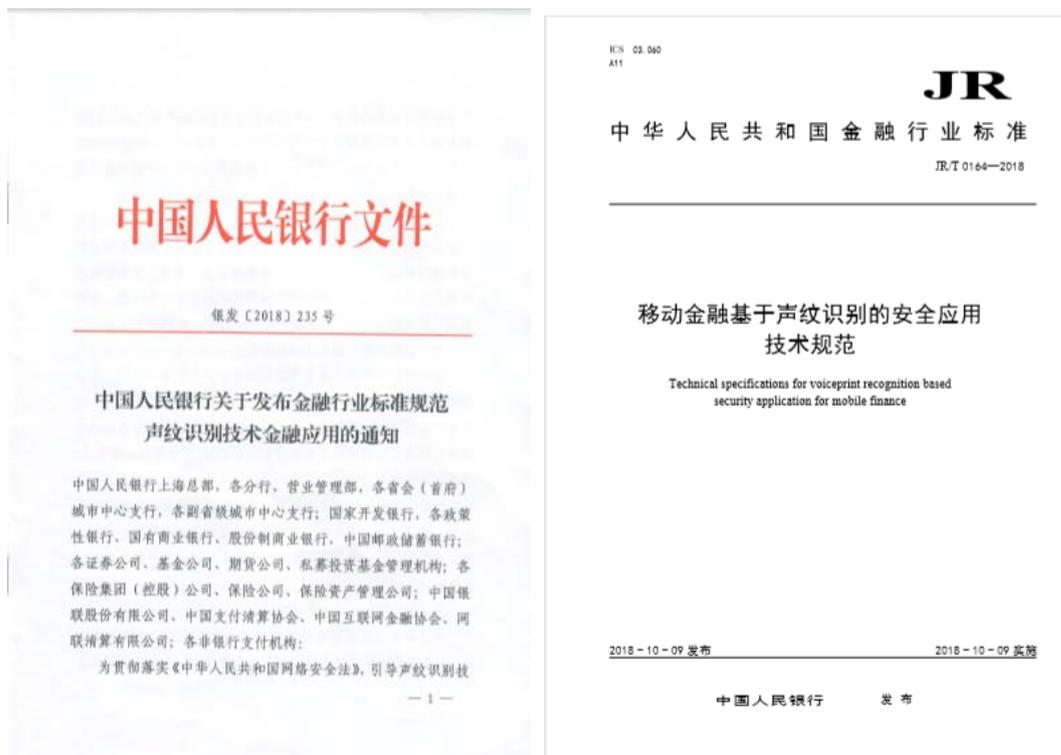
5.2. 金融应用标准化

2016年7月11日，由中国建设银行、清华大学、北京得意音通技术有限责任公司负责起草的《手机银行中基于声纹识别的增强安全技术规范》，经专家组评审后由金标委立项（项目编号 FISP - 2016 - 005）。

2016年7月开始，在央行科技司主导下，发起单位联合工、农、中、交等国内各大银行，国家级测评机构及第三方支付平台等共同对标准报批稿进行了为期一个半月的集中修订完善，并将标准适用范围从“手机银行”扩大到“移动金融”。

2018年8月18日，由央行科技司召集专家评审会，来自中央网信办、公安部、国家信息中心、国家信息安全研究中心、中国银联、中国科学院、中国社会科学院、清华大学等单位的院士专家一致认为，“与其他生物特征识别技术相比，声纹识别技术在个人隐私保护、身份认证强度等方面有一定的优势”，“具有较高的可靠性、安全性、便捷性，能够满足移动金融服务应用需求”。

2018年10月9日，《移动金融基于声纹识别的安全应用技术规范（标准编号：JR/T 0164-2018）》由中国人民银行正式发布，即日起实施。与此同时，《关于发布金融行业标准规范声纹识别技术金融应用的通知（文件编号：0001-2018-S-000-005057）》随该标准的印刷版一同下发给全国各大银行及各类金融机构。



图表 5-3 《移动金融基于声纹识别的安全应用技术规范》

该标准由中国建设银行、清华大学、北京得意音通技术有限责任公司发起，从前期调研、立项论证、标准起草到征求意见、修订和报批，前后历时两年多，通过了 48 位金标委委员的投票表决，和 11 位院士专家的集中评审，历经“四稿三审”才正式由中国人民银行发布，充分体现了技术新、起点高、过程严、范围广、自主性强等特点。

中国人民银行副行长、全国金融标准化技术委员会主任委员范一飞指出，这是“我国金融行业生物识别的第一个技术标准，标志着以声纹识别为代表的生物识别应用进入崭新的历史发展阶段”。

5.3. 正在进行的标准化

2018 年 5 月 14 日，全国安全防范报警系统标准化技术委员会（简称安标委，秘书处设在公安部第一研究所）下设的人体生物特征应用分委员会（SAC/TC100/SC2），投票通过了声纹识别标准化体系建设 12 项标准中的 3 项，进入起草阶段，它们是：

- 《声纹数据采集的技术要求》

- 《声纹数据质量评价标准》
- 《声纹数据建库要求》

这三大标准制订是国家重点研发计划“国家质量基础的共性技术研究与应用（NQI）”专项——“智能语音产品符合性测试技术研究”（2017YFF0210901）的子任务，由得意音通与清华大学牵头，对规范我国未来声纹身份认证具有重要意义。

2018年5月25日，两项应用类标准由全国信息技术标准化技术委员会（简称信标委，秘书处设在中国电子技术标准化研究院）予以立项，进入标准起草阶段。它们是：

- 《信息技术 移动设备生物特征识别 第5部分：声纹》
- 《信息技术 生物特征数据交换格式 第13部分：声纹数据》

这两项标准是前述三大标准的延伸，主要规范了数据交换中的格式要求，及在互联网金融等重点领域的应用要求。

此外，由中国信息通信研究院牵头的中国人工智能产业发展联盟（AIIA）、由蚂蚁金服牵头的互联网金融身份认证联盟（IFAA），这两大行业联盟的声纹识别团体标准也已提上议事日程。一个覆盖技术、应用、数据、评测等的声纹标准化综合体系已经成型。

6. 技术挑战与发展趋势

6.1. 技术挑战

由于声音信号的复杂性，声纹识别在实际应用时仍然面对很多挑战，根据识别任务的不同，挑战也会稍有不同。比如：

身份认证（声纹确认）：假冒攻击、真实意图检测、低语识别等；

反欺诈（声纹辨认）：多说话人、短语音、语音掩盖等；

文本无关：（录音重放的）假冒攻击、短语音；

文本相关：（录音重放的）假冒攻击；

文本提示：（录音拼接重放）假冒攻击；

共性：背景噪音、跨信道、时变（年龄变化）、身体状况、（语速、音量、情感等）发音方式。

下面分别叙述。

1) 背景噪音

声纹识别的实际使用中，不可避免的会带入环境的噪声，如办公室大厅的人声、马路边的汽车声、超市菜市场的嘈杂声、车上的引擎声和音乐声等。此类噪声会扰乱语音信号，对声纹识别性能造成影响。由于噪声的多样性和未知性，完全消除噪声的影响十分困难。目前，一些研究者语音增强的方法。一些方法在信号层对噪声进行估计，并尝试对其进行消除，在稳定噪声信号下得到了不错的效果，但对于非稳定信号效果不佳。另有一些研究者尝试在特征域、模型域和分数域，对噪声造成的影响进行抑制。还有一些研究者通过数据增强的方法，通过模拟方法在训练数据中加入噪声。这些方法均取得了一定效果，但在实际的复杂多变的噪声环境下，噪声问题仍然是一大挑战。

2) 跨信道

声音需要通过复杂的信道环境，才能最终传输到系统当中。信道即信号传输的通道，包含用户嘴唇处形成的声音信号，到系统接受到的声音信号之间的所有过程。常见的信道有电话信道、网络信道等。此外，录音设备的不同，也是信道差异的一个重要来源。信道会对语音信号产生畸变，影响语音信号的听感。此外，信道还会夹杂信道噪声，扰乱语音信号。因此，同一个说话人在不同信道下的声

音，会有很大的差别。这种差别影响了声纹识别系统的识别性能。如何在跨信道情况下进行声纹识别任务，是当前一个很大的挑战。

3) 短语音

短语音是指系统的识别性能对语音长度的依赖性。较长的语音会达到更高的精准度，但显然，过长的语音会影响用户体验。并且在一些特定场景下，比如司法应用中，系统只能收集到有限长度的语音。因此，如何在较短语音长度的情况下，提高系统的识别性能，也是一个研究方向。

4) 假冒攻击

声纹识别作为一种身份认证算法，经常应用于安全领域。此时，必然存在一些攻击者，试图愚弄声纹识别系统，以绕过安全验证，达成其攻击目的。声纹识别的防假冒攻击即指声纹识别系统拒绝非真实说话人的能力。这些尝试进入系统的声音，可能是由人类模仿发声的，也可能是机器伪造的，比如通过语音合成、声音转换以及录音重放技术，产生出和真实说话人相近的声音，尝试进入系统。目前，语音合成技术和声音转换技术达到了一个新的高度，计算机已经可以合成流畅的、十分逼真的、甚至是人类难以辨别真伪的语音，此类语音对声纹识别系统有很大的威胁性。录音重放攻击则直接录制目标说话人的真实声音，其具有很高的还原度。目前，已有研究者对于语音合成、声音转换以及录音重放的攻击进行了研究，并进行了防御对策的研究，取得了初步效果。但由于真实场景下攻击方式未知切多样，如何阻止这些假冒语音通过系统，仍然是一个重要的研究方向。

5) 真实意图

声纹识别技术经常应用于安全领域。除了假冒攻击外，攻击者可能使用胁迫的方式，要求真实说话人进行验证。此外，也存在真实说话人在醉酒、说梦话甚至在无意间，通过声音进行了验证。此时，真实意图检测就显得非常重要。真实意图检测，即检测验证者在进行验证时，是否具有真实的被验证的意愿。若发现用户并非主动自愿的进行验证，则系统将发出警告，并进行恰当措施。然而，意图是一个抽象概念，即使是人类，也很难非常准确的通过语音来了解对方的意图，其更多的是一种抽象的感觉。同时，除了语音内容所表达的显式的意图外，更多的隐含的意图特征分散的隐藏在声调、语速等更多细节中，因此检测这类隐藏将十分困难。因此，如何通过语音，对说话人的真实意图进行检测，是一个非常具

有挑战的研究方向。

6) 时变（年龄变化）

人的身体机能会随着年龄增长而变化，而其声音也会随之而发生变换。对于同一个人间隔较长的两次录音，其声音也会有一些不同。这种不同会削减两个语音的相似性，进而导致声纹识别系统的性能下降。目前，有研究者通过自适应的方式，不断使用新的数据自动更新模型。但此方式无法解决长期间隔的问题。另有研究者尝试提取时变鲁棒的特征，或构建时变鲁棒的模型，取得了一定进展，但如何在人的年龄变化的情况下进行鲁棒的声纹识别，仍然是一大挑战。

7) 身体状况

人的声音会随着身体状况发生变化。生活中，难免会碰上感冒、发烧、鼻塞、嗓子发哑的时候，此时，因身体状况的变化，发声器官本身发生了改变，因此人所发出的声音特质也会随之改变。比如当鼻塞时，鼻腔通道关闭，声音会发闷；嗓子发哑时，声带激励的特性会发生变化。由于身体状况变化直接导致发声器官变化，影响声纹识别的准确性，因此，如何在身体状况发生变化时，进行鲁棒的声纹识别，是一项重要的挑战。

8) 发音方式

语音信号，不仅与说话人和说话内容相关，也与说话人的发音方式相关。不同的发音方式，如语速快慢、音量大小、语气变化等，会带来不同的听觉效果。由于发音方式是发音器官在发音时的一种行为特性，此类信息在语音信号中占有较大比重，即使在语音内容相同时也会对语音信号的频率成分造成影响，干扰声纹识别的准确判断。因此，如何在发音方式发生变化时，进行鲁棒的声纹识别，是一个具有挑战的研究方向。

9) 低语

声纹识别需要通过发声才能进行认证，然而存在大量场合（如开会、上课、有人在休息时等）需要保持安静，不能大声说话。此时，若能够通过低语进行声纹识别，则可很大程度上提高声纹识别的应用场景，提高用户体验。低语分为两种，一种为声带振动的低声说话；另一种为耳语，即在声带不振动的情况下进行说话。由于用户在低语情况下进行发声，其发音习惯与平常不同，因此发声将造成变化。对于耳语情况，由于其声带不振动，相对正常发音，丢失了许多说话人

信息，这将给声纹识别造成困难。此外，由于耳语发音的特殊性，其对信道、说话人身体状态、心理的变化等也更加敏感。因此，如何在低语情况下，进行鲁棒的声纹识别，是当前声纹识别的一大挑战。

10) 语音掩盖

在特殊情况下，有些人不想自己的声音被别人认出，就会对声音进行掩盖，如使用假声说话，使用变声器等。此类情况即成为语音掩盖。然而，一些应用场景需要对经过掩盖的声音进行声纹识别，判断语音背后说话人的真实身份。由于在语音掩盖情况下，说话人会刻意改变其声道形状、发声习惯等，同时会故意增加声音的变化性，这使得系统需要寻找一种稳定的、难以按照人类主观意愿进行改变的声音特征来进行识别。因此，如何做语音掩盖情况下，对上说话人身份进行识别，是一个研究挑战。

11) 多说话人

在实际应用场景下，可能会出现多个说话人轮流或者同时说话的情况，比如呼叫中心的客服对话、会议录音等。此时，用户希望将语音中多个说话人进行分离，以方便进一步处理。由于当前声纹识别技术，其准确性往往与语音时长相关，较长的语音才能进行更准确的识别。而当同一语音包含多说话人的情况下，需要对不同说话人之间的边界进行准确的界定，这就需要对短时间内说话人的变化进行检测。此外，多人说话时常常出现同时发声的情况，此时，如何对同时发声进行识别、处理、分离，也需要进一步的研究。因此，如何能在同一段语音中，对有多个说话人的身份进行识别，进而对语音进行分割，是当前一个研究方向。

6.2. 发展趋势

1) “数字身份”兴起

生物识别技术正逐渐成熟并进入大规模应用阶段。随着 3D 传感器的快速普及、多种生物特征的融合，每个设备都能更聪明地“看”和“听”。生物识别和活体技术也将重塑身份识别和认证，数字身份将成为人的第二张身份证。

2) “声纹+”多模态融合

由于不同场景对生物特征的适应性各有不同，基于多模态的解决方案代表了一个新兴趋势，多生物特征识别技术融合已经成为必然。国际标准化组织（ISO）

和国际电工委员会（IEC）已经联合公布了《信息技术—生物特征—多模态和其他多生物特征融合》（ISO/IECTR24722：2007），该方案能融合多种生物指令，以保证在一种生物特征失真的情况下，仍能顺利识别。此外将多种识别方式叠加起来，攻击成本会急剧上升，从而带来安全性上升。这有助于降低金融诈欺所造成的损失，确保用户信息或数据安全。

3) 5G 催生全新应用场景

第五代移动通信技术将使移动带宽大幅度增强，提供近百倍于 4G 的峰值速率，促进基于 4K/8K 超高清视频、AR/VR 等沉浸式交互模式的逐步成熟。连接能力将增强至百亿级，带来海量的机器类通信及连接的深度融合。网络向云化、软件化演进，网络可切片成多个相互独立、平行的虚拟子网络，为不同应用提供虚拟专属网络，加上高可靠、低时延、大容量的网络能力，将使车联网、物联网、工业互联网等领域获得全新的技术赋能。语音作为最自然的人机交互方式，将迎来更多成长机会和更大的发展空间。

附一：图表索引

| | | |
|---------|-----------------------------------|----|
| 图表 1-1 | 2007—2021 年全球生物识别技术行业市场规模与预测..... | 2 |
| 图表 1-2 | 2002—2021 年中国生物识别技术行业市场规模与预测..... | 2 |
| 图表 1-3 | 移动支付用户对生物特征识别技术认知程度分析..... | 3 |
| 图表 1-4 | 移动金融中远程身份认证技术安全性比较..... | 4 |
| 图表 2-1 | 生物特征识别技术特性对比..... | 7 |
| 图表 2-2 | 声纹识别技术发展简史..... | 10 |
| 图表 2-3 | 国内声纹相关专利申请数量宏观态势..... | 11 |
| 图表 2-4 | 国内声纹相关专利类型分布..... | 11 |
| 图表 2-5 | 国内声纹相关专利法律状态..... | 12 |
| 图表 3-1 | 中国声纹识别相关厂商成立时间轴..... | 13 |
| 图表 3-2 | 中国移动金融市场上的主要声纹识别厂商分布..... | 14 |
| 图表 3-3 | 各声纹技术相关厂商技术来源..... | 15 |
| 图表 3-4 | 中国声纹识别相关企业最新融资状况..... | 15 |
| 图表 4-1 | 声纹识别技术分支-应用场景分布图..... | 17 |
| 图表 4-2 | 中国声纹社保试点分布图..... | 19 |
| 图表 4-3 | 中国建设银行声纹注册用户模型数增长变化..... | 20 |
| 图表 4-4 | 建行手机银行 APP 声纹建模/更新界面..... | 21 |
| 图表 4-5 | 建行手机银行 APP 用户双重身份认证登录..... | 22 |
| 图表 4-6 | 建行手机银行转账声纹验证过程..... | 23 |
| 图表 4-7 | 建行手机银行声纹取款业务流程示意图..... | 23 |
| 图表 4-8 | 贵阳银行手机银行理财产品交易声纹验证..... | 24 |
| 图表 4-9 | 贵阳银行手机银行声纹注册界面..... | 24 |
| 图表 4-10 | 西安银行手机银行声纹登录场景..... | 25 |
| 图表 4-11 | 浦发银行手机银行声纹登录..... | 26 |
| 图表 4-12 | 浦发银行手机银行声纹注册界面..... | 26 |
| 图表 5-1 | 《自动声纹识别（说话人识别）技术规范》..... | 27 |
| 图表 5-2 | 《安防声纹确认应用算法技术要求和测试方法》..... | 28 |
| 图表 5-3 | 《移动金融基于声纹识别的安全应用技术规范》..... | 30 |

附二：参考文献

- [1] 郑方、李蓝天等，《声纹识别技术及其应用现状》，《信息安全研究》2016年第1期
- [2] 陈孝良，《声纹识别技术的现状、局限与趋势》，雷锋网“硬创公开课”，2016年11月
- [3] 中国电子技术标准化研究院，《生物特征识别白皮书（2017年版）》，2017年7月
- [4] 电子商务与电子支付国家工程实验室，《“声纹识别在金融领域的应用”系列报告》，2017年12月
- [5] 世界银行，《世界银行：数字识别技术概览（2018英文版）》，2018
- [6] 李智勇，《语音交互的进阶之路：层级、技术与颠覆性》，CSDN《2018中国人工智能产业路线图 V2.0》，2019
- [7] 李伟，《推动生物识别技术在金融领域安全规范应用》，《中国信息安全》2019年第2期
- [8] 郑方、程星亮，《声纹识别：走出实验室，迈向产业化》，《中国信息安全》2019年第2期