

本文将谈谈在智能音箱、智能机器人等此类以语音为主要交互方式的智能硬件产品中，除了提升语音识别准确率、语义理解效果外，怎样使用声纹识别来进一步提升用户体验，让智能产品更加智能。



一、背景

当前智能硬件产品中，最耀眼的莫过于百箱大战的智能音箱、百“机”争鸣的智能机器人，这些智能语音产品已逐渐走进百姓的视线中。

在智能音箱市场中，且不说国外的 Amazon Echo 和 Google Home，仅在国内，去年双十一，天猫精灵 99 元跳楼价卖了一百万台，还有铺天盖地而来的小爱同学、叮咚音箱、小雅同学、出门问问等。

今年，百度推出比天猫精灵还低 10 元的小度智能音箱，仅卖 89 元，烧钱大战一个比一个狠。在智能机器人市场也同样打得火热，这类智能音箱和智能机器人最明显的特点，就是采用了语音作为全新的交互方式，力图打造更接近于人与人交流的方式。

然而，大部分智能语音产品只能识别出说话的内容，对于相同的问题都会给出相同的反馈信息。虽然交互方式是智能的，但交流方式仍然不够智能。

所谓交流，即有明确对象，和不同对象交流应有不同的回应，闻声即可识人，这才是更智能的体验。声纹识别，正是可以实现这种闻声识人的技术手段。

本文将谈谈在智能音箱、智能机器人等此类以语音为主要交互方式的智能硬件产品中，除了提升语音识别准确率、语义理解效果外，怎样使用声纹识别来进一步提升用户体验，让智能产品更加智能。

二、应用场景

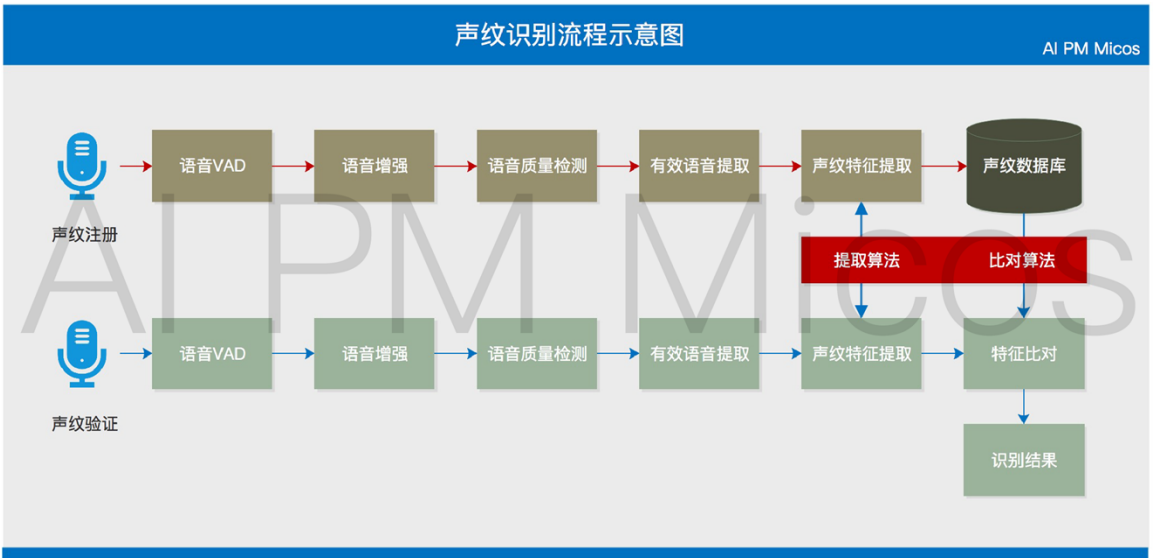
场景是需求的灵魂，抛开场景谈需求都是耍流氓！

我们先对场景分分类，声纹识别作为身份验证的一种手段，根据不同的安全性要求，我们将应用场景分非关键应用场景和关键应用场景。

非关键应用场景：即那些对安全性要求不高，即使误识也不会产生明显不良影响，但对用户体验要求较高的场合，如：猜猜说话人是谁、个性化的服务推荐（歌曲推荐、餐厅推荐等）；关键应用场景：这类应用场景有明显的安全性要求，不能容忍有误识情况，误识会带来明显的不良影响，如：一些重要的权限控制（家电开关控制等）、声音购物支付等，这类型场景会牺牲一部分用户体验以保证安全性，毕竟鱼和熊掌不可兼得！

三、应用设计

声纹识别在应用上分为注册和验证两个流程，如下图所示：



在智能音箱和智能机器人（家庭陪伴机器人）的产品中，声纹识别其实就是 1: N 的应用（N 的值一般小于 10），音箱和机器人一般都会有一个名字，即唤醒词，就如同人的名字一样，以下设计我们暂且给智能硬件起个名字叫“你好同学”。

1. 声纹注册流程设计

声纹注册是开启闻声识人的第一步，注册者先说几句话，系统自动将其中的声纹特征提取出来作为说话人身份的 ID，这个过程如同像陌生人做自我介绍一样，介绍完大家就认识你了。

虽然声纹注册流程必不可少，但产品设计上却希望这个流程越简单越好，最好是不需要注册这个环节（当然这是不可能的!!!）。

在做注册流程设计前，我们先搞清几个问题：

（1）注册入口在哪里？

目前的智能语音产品在交互方式上，除了用语音，还会分为带显示屏和不带显示屏，即有部分带有视觉交互。

所以，注册入口可分为三种：

直接在设备上使用语音方式作为注册入口，即直接对设备说：你好同学，我要注册声纹；在设备配套的 APP 上设置注册入口，适合无屏的设备；在设备自带的屏幕上设置注册入口，类似 APP 方式。

（2）在哪里采集注册录音？

对于注册入口在 APP 上的情况，我们需要搞清楚录音采集是在设备上，还是在手机上。

一般来说，无论注册入口在哪里，最终的验证入口都是在设备上，为了避免在不同设备上录音效果不一样，造成验证准确率下降（我们叫这种现象叫信道失配，即注册信道与验证信道不一致而带来的准确率下降的现象），我们会直接在设备上进行注册录音的采集。

当然，你非得在 APP 上录音，也是可以的，呵呵！

（3）非关键应用和关键应用是否需要单独注册？

理论上来说，注册语音越长，识别效果越好，但也需要考虑到用户体验。所以，我们可以将注册流程设计为基础注册+加强注册，其中基础注册满足非关键应用场景，加强注册主要是考虑到关键应用场景的安全性问题。

(4) 声纹注册上还有哪些需考虑的点？

智能音箱和机器人一般都是远场语音应用，交互距离最大可达 5 米，同时也覆盖了近场语音，所以在说话人离设备的距离不同时，采集的语音质量也会存在差异。对语音识别来说可能不会有太多的影响，但对于声纹识别来说，会造成声纹特征的损失。所以在注册时还需要考虑不同距离，比如：0.5 米、3 米、5 米等。

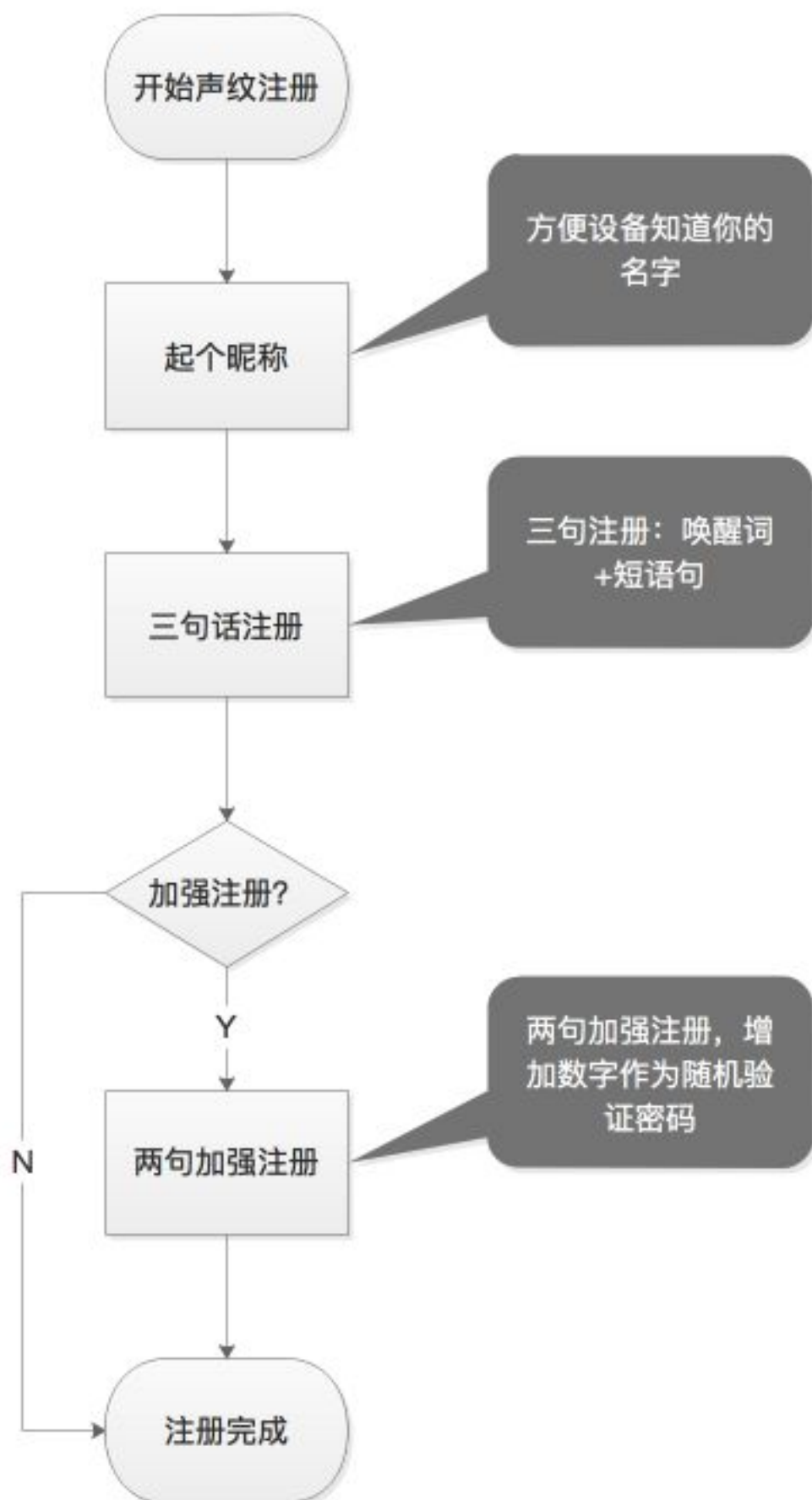
另外，还需要考虑说话的音量、语速等因素，相信随着识别算法的提升，这些因素都不会成为产品设计的考虑点。

(5) 声纹注册有哪些表现形式？

在注册的表现形式上，可分为两种：

引导式注册：根据界面或语音提示，引导用户一步一步进行注册，这是最为常用的方式，也是目前最合适的方式。**无感知注册：**即不需要指定用户执行注册流程即可完成声纹注册，打破常规的声纹注册-声纹验证的应用流程，在使用过程中自动完成声纹注册，直接实现“听声辨人”，最大限度提升用户体验。我们下次再来详细讲一下这种无感知注册。

2. 声纹注册流程



3. 原型设计参考

以带屏设备上或 APP 上注册为例，原型仅供参考，在真实项目中仍需要考虑更多的交互细节。对于注册内容，一般建议使用“唤醒词+常用短语”的方式。当然，如果追求更简洁的注册方式，也可以只读两次唤醒词。



4. 声纹验证流程设计

在声纹验证环节，一般通过纯语音交互来完成，我们按非关键应用场景和关键应用场景来设计。

(1) 非关键应用场景验证

最常用的自我身份验证：对设备说出“你好同学，猜猜我是谁”或“你好同学，我是谁”，设备根据声纹识别结果回应说话人，比如：设备回应：哎哟，你就那玉树临风、风流倜傥、迷死万千少女兼大妈的彭鱼宴。

非关键应用场景有非常多的玩法，主要围绕不同身份的个性化推荐来设计，具体大家可以根据实际产品及场景需求来思考。

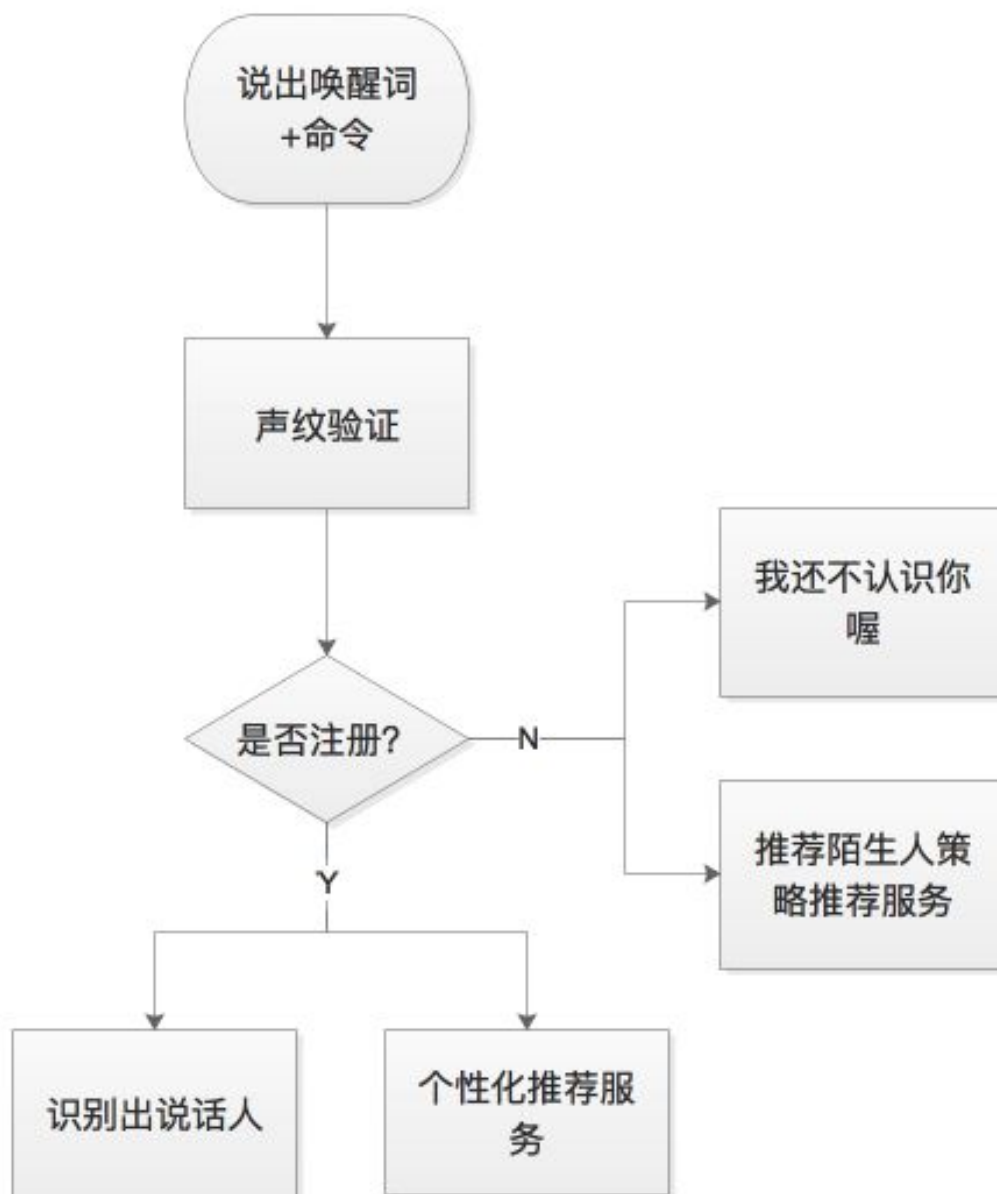
另外，对于带屏的设备，一般都会具备摄像头，可以实现人脸识别功能。有些厂家可能会将人脸和声纹结合起来做身份验证，这种多维度验证确实能提高安全性，但建议只在关键应用场景下使用两者作为验证手段，在非关键应用场景中只使用声纹识别。毕竟如果每次问机器人我是谁时，还要自己找到摄像头正面看一会才能识别出身份，这样的体验非常不好。

(2) 关键应用场景验证

此场景中，需要考虑二次身份核验，即在唤醒时已做一次身份识别，在说话人做出关键操作确定后，比如：确定购买商品，需要进行身份核验。

由于是纯语音交互，考虑到安全性，可采用读 4 位随机数字或唤醒词的方式来验证，此种情况下声纹识别必须具备活体检测功能，以防止通过录音的方式来假冒说话人声音。

(3) 验证流程设计（实际流程远比这个复杂）



四、未来应用

虽然目前的智能硬件仍在不断完美语音识别和语义理解的效果，然而，我们相信增加声纹识别必定会大大提升智能设备的使用体验，同时亦可增加情绪识别、性别识别、年龄识别，让机器更懂你。

正如电影“her”中的萨曼莎，语音交互水平已达到了强人工智能水平，比与人类交互效果更好。因为 her 可以在一秒内检索完主人所提问题关联到相关知识信息，以更好给出回

答，这是人类所不能达到的。要达到人类期望值中的语音交互效果，就必须让设备能够：听得清、听得懂、学得会、说得出。

期待 “her” 的出现!!!

作者：Micos，在人工智能浪潮中推波助澜的产品经理，致力于用智能语音实现人与机器最自然的交互方式。

题图来自 Unsplash，基于 CC0 协议