2017

声付 ShengFu

声付的实现与使用场景

毛一鸣

目录

1.声付的想法与实现基础

1.1 支付宝与空付

大家知道诸如手机银行、支付宝、微信支付这类移动支付工具和我们的钱包紧密相连。因此为了提高支付安全性，支付厂商们不断改进支付手段，从最早的字符密码、手势密码，逐渐升级为依赖人体生理特征的指纹支付、虹膜支付、刷脸支付等更安全的支付技术。但是这些支付手段都有或多或少的局限性。因此为了更方便用户的使用，支付宝研发一项新的支付技术——空付。它可以使用你拥有的任意一个生理标记，如纹身、疤痕等来实现支付。这样以后使用支付宝时，即使你没有携带手机也可以实现安全支付。

空付的出现，从根本上的改变了传统的无现金支付解决方案。对于一个传统的支付流程（图 1），他需要用户使用手机，并对这次交易使用各种方式授权，例如刷脸，例如手势密码等等。

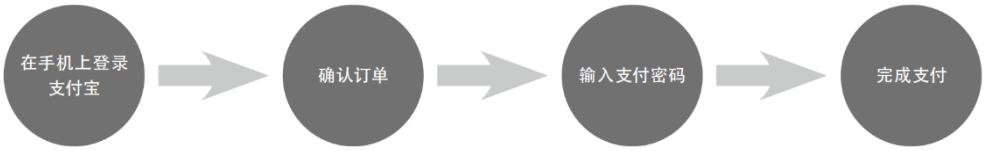

图 1

但是空付的出现则不一样，空付顾名思义就是两手空空也能付钱，正如上所说，你将任意一个生理标记作为支付的证明。那么你肯定是随身携带这些标记的，那么只要商家扫描你的空付证明，就可以完成这笔交易（图 2）。

图 2

1.2 空付的技术条件

空付技术包含两个部分，首先是APR（Augmented Pay Reality，即增强支付现实技术），它是一种先进的现实技术。在使用空付之前，支付系统首先需要对被拍摄对象进行检测分析，提取特征。它的原理实际上和指纹支付一样，只不过空付的APR系统可以扫描你身体的任意部位（也可以是任意物体，如你家的小猫、小狗等）的特征点。APR系统扫描完指定的特征点后，接下来就是将这个特征点和你的支付宝账户关联。同时可以对特定支付的场景（如仅限于超市、便利店）及支付限额（如限制该方式只能用于小额支付）进行设定。完成上述数据的采集后，它会将特征点转换为数字化数据（确保这些数据可以被支付扫描系统如超市的扫描枪识别），最后APR会将数据保存在支付宝系统云服务器上。

这样当我们需要使用空付作为自己的支付手段时，此时空付的另一个重要部分

IRS（Information Recall Secure）即信息回溯保障系统就出现了。在配备IRS的收款设备需要收款时，此时支付宝用户只要展示前面预先设定的特征点，IRS系统就会根据APR技术解析后的信息，然后自动连接到支付宝系统云服务器上，通过和服务器预先保存的数据进行比对，如果一致则激活指定的支付宝账户完成付款，从而实现“空付”。

1.3 声付的想法

在现阶段，支付宝已经用它强大的技术水平实现了利用人身体特征作为支付凭证，但是除了人体特征的图片之外，人还有一种极易获取的生理特征——声音。作为人类最基本的几个特征：指纹，声音，虹膜。其中指纹的支付已经被广泛使用，而虹膜的识别需要更多的设备支持，而如果用于支付凭证，想必用户的体验也不会很好（需要有设备近距离扫描你的眼球），而声音则完全不同，不论是获取难度，还是处理速度，现在的声音识别的足以胜任，而且，对于商家来说，一台联网的录音设备，就可以是一个完整的声付支持设备，这相比于指纹、空付来说，显得更为快捷与方便。

2.声付的实现技术

2.1 声纹提取技术

声纹提取是一项根据语音波形中反映说话人生理和行为特征的语音参数, 自动识别说话人身份的技术。它的基本原理是通过分析人的发声和听觉, 为每个人构造一个独一无二的数学模型, 由计算机对模型和实际输入的语音进行精确匹配, 根据匹配结果辨认出说话人是谁。首先对鉴别对象的声音进行采样, 即输入语音信号, 再对采样数据进行滤波等处理, 而在声纹识别过程中最主要的两部分内容是特征提取和模式匹配。特征提取, 就是从声音中选取唯一表现说话人身份的有效且稳定可靠的特征;模式匹配就是对训练和鉴别时的特征模式做相似性匹配（图 3）。

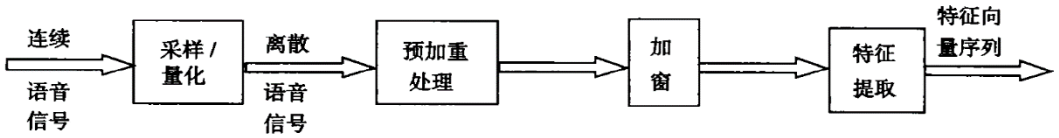

图 3 特征提取

2.2 声纹提取之于声付系统

所以，基于如上的考虑，对于声付的想法来说，想要有效的获取用户的声纹数据，首先需要建立起有效的声纹数据库，这一部分必然是存在与用户的客户端上。首先的想法就是让系统可以接入 Siri，作为 IOS 的智能语音控制系统，Siri 在日常的工作中必然的收集了用户的语音数据，那么，如果能够读出这些数据，将可以有效的分析建模，计算后获得有效的模型。如果无法通过 Siri 获取用户的声音数据，那么声付系

统将进行自己的用户语音数据采集，他会提供用户一些文本片段，让用户朗读出来，然后声付系统会录音并在后台进行处理后将获得的声纹数据上传至服务器，同时，在用户同意的情况下，还可以通过获取用户日常的语音聊天，或者电话录音，来获取用户的声纹数据，以求提高采样率，获得更加适配用户个人声纹的特征模型。

2.3 声纹识别技术

声纹识别是一种自动识别说话人的过程。它是人体个性特征识别中的一个重要分支，它是根据语音波形中反映说话人生理和行为特征的语音参数自动识别说话人身份的技术。用户在使用声纹识别系统时，需要像系统提供一段语音，根据所需语音的需要，可以将声纹识别系统分为：与文本有关和与文本无关的声纹识别系统。就目前的语音环境下，声付系统应当优先考虑的是使用与文本有关的声纹识别，因为这种识别需要用户按照规定的内容发音，根据此建立精确的模型，训练与测试语料一样。因此这种识别方式的识别效果非常好，但需要用户配合，如果发音内容与规定的内容不符，则无法正确的识别用户。

2.4 声纹识别技术之于声付系统

对于声付系统来说，声纹识别技术因被应用于商家的客户端上，构建一个文本无关的声纹识别并不是十分容易，这需要大量的对于用户个人的采样，但是作为一个便捷的支付系统，大量的采样显得违背了初衷，所以声付系统当考虑的是进行文本有关的分析。综合上述声纹提取方法，让用户朗读特定的文本特征片段之后，进行声纹提取，而后再生成 10-15 句简短的文本，让用户再次朗读确认后，当用户外出进行声付时，商家出示 10-15 句本文中的 1-2 句。让用户进行朗读确认。目前的文本相关声纹识别的技术非常快，只要 0.3-0.5 秒即可完成从确认到支付的整个流程。

3.声付的场景综合考虑

3.1 嘈杂背景环境

对于声付系统来说，最大的考验莫过于在嘈杂的背景声之下对于用户声纹的识别。在很强的背景声下，用户的语音会被很多外界的声音干扰，如果背景也存在其他人的语音时，这种混合的声音必然很容易被混淆而导致声音识别的失败。

作为声付系统，其首先自然是希望处于较为安静的环境状态。但依旧无法排除会有嘈杂的干扰，例如在公园内某个小摊位上进行支付。为此，我们提出的解决方案是：

一、我们可以在商家端的声纹识别系统中加入隔离噪音的插件，这种插件可以是物理性质的，例如减少收音域，使用隔音材料等等；也可以是软件性质的，例如，通过 10 秒或者 20 秒一次的平均采样，对该地区的背景声进行分析，当用户进行声付时，利用软件剔除这一部分噪声。

二、对于突发性的背景噪声，例如当用户进行声付时，朋友也在和他说话，这样的情况会导致 2 人的语音都被进行分析，针对这种情况，文本有关识别就可以发挥他的用处，我们可以先对录入的去背景语音进行再次分析，提取出符合商家出示的文本的语音片段，而后在交付服务器进行声纹分析。

这些技术都是存在于商家的服务端内，并不需要服务器本身承担除杂和选取的语音的操作，因此并不影响服务器的并发处理速度，用户可以同样获得很好的声付体验。

3.2 不安全环境

对于声付系统来说，另一个考验就是安全性，对于不法分子来说，获取用户的语音数据相对来说显得容易很多，因此声付系统能否克服这类不法方式就显得极为重要。

大体来说，获取用户的声纹欺骗支付系统的方式有两种，第一种是针对声纹进行模拟的攻击，即通过获取用户的语音，利用软件模拟用户的语音，朗读商家文本进行欺骗，第二种是针对文本相关声纹识别技术的攻击，通过录取用户所有可能朗读的 10-15 句语音片段，通过重放的方式进行欺骗，针对如上两种手段，目前依旧没有十分好的解决方案，但是依旧可以做到一定的保护：

一、针对第一种欺骗，我们可以使用自行研发的声纹识别技术，目前大部分论文中提到的声纹识别技术都是公开算法的，只要不法分子了解这种算法，只要能够获取用户的语音，便可以获取整个用户的声纹模型，然后进行逆运算，通过软件模拟出相同声纹特征的语音片段。因此通过自行研发识别技术，加大算法保密手段，可以做到让不法分子无法破解模型，也就无法进行声纹模拟攻击。

二、针对第二种欺骗，它的难度非常大，因为我们是从 10-15 个文本中选取 1-2 个让用户朗读，所以不法分子需要录音到至少 15 次交易语音才有可能完成攻击，但是这种可能性依旧是存在的。所以我们提供的解决方案时，我们会希望用户定期的录入新的文本片段的语音，以加大不法分子的攻击难度，这个周期可以是 1-2 个月。同时，我们可以研发对于录音语音的识别技术，因为很显然的，当用户直接朗读文本时，这段语音经历过了“空气传声-固体传声（录音设备）”的过程，而当不法分子播放录音时，这段语音经历的是“固体传声（录音设备）-空气传声-固体传声”。声音的音调会发生变化，所以应当可以研究算法，去捕获这个变化，以完成通过录音的方式进行文本相关声纹识别的攻击的防御。

4.声付之于未来

通过上文，我们可以看出，声付存在极佳的前景，他可以做到比空付来的更加方便，毕竟语音才是最快捷的信息交流方式。但是声付同样存在这一些技术性问题，对于一些不安全的因素还是没有完全的防御方法。但是，我认为，通过如上的分析与提出的解决方案，声纹支付已经是一个小额消费支付的合理解决方案。

参考

周雷 上海师范大学 《基于声纹识别的说话人身份确认方法研究》

柳絮飞 《技术宅》 《不带手机 支付宝玩“空付”》

张万里，刘桥 贵州大学电子科学系 《Mel 频率倒谱系数提取及其在声纹识别中的应用》