# 带反馈的连续声纹认证技术

## 技术背景

生物身份认证通过用户特有的生物特征进行身份确认或辨认，在现代社会中具有重要意义。声纹身份认证技术是各种生物身份认证中的重要一员。与其它认证方式相比，声纹认证具有方便性、便捷性的特点，特别适合远程认证，在移动支付时代具有重要意义。所谓声纹认证，是指用声音来判断说话人身份的方法。典型的场景是：用户首先念一段话进行注册；在认证阶段，用户同样念一段话（内容可相同，也可不同），对比该段认证语音和注册语音的异同，如果这两段语音在发音人特性上足够相似。即可通过认证。

当前声纹认证技术都可称作“无所馈片段认证”。在这一认证方式中，用户按要求给出一段发音，在发音过程中系统无反馈；发音完成后，系统对该声音片段进行整体评价，并给出结果。本发明提出一种“带反馈连续认证”方式。在这种新方式下，用户在认证的过程中可即时得到自己语音的匹配分值，并可依该分值调整说话方式，以配合系统完成认证。这一新方式将带来更好的用户体验。

## 发明内容

传统无反馈片段认证模式步骤如下：

1. 系统提示用户发音以进行身份认证
2. 用户按系统提示发音
3. 用户发音完毕后，系统给出一个拒绝或接受的判定，同时也可能给出判定的分数

图1给出这种认证方式的示意图。



图1：传统无反馈片段认证方式；（左）发起认证，用户按住下方按钮发音；（右）发音结束后，系统给出认证结果

这种无反馈片段式认证方式存在的一个弊端是认证过程对用户是不透明的，即用户在这一过程中不能得到任何反馈。对非法用户来说，这种方式可减少系统的信息暴露，从而减小被非法用户闯入的风险；但对合法用户来说，这种方式会极大提高被错误拒绝的可能。这是因为用户在认证过程中缺乏反馈机制，无法通过调整发音方式来配合系统提高认证分数。

类比于人脸识别，这相当于举起相机随便拍张照片让系统确认。这种缺乏反馈的认证方式会极大提高系统的错误拒绝率，使合法用户因拍照姿势、光照条件的变化无法通过验证。错误拒绝的增加会极大降低用户使用体验。为解决这一问题，人脸识别系统会给出摄像头拍出的实时影像作为反馈，允许用户调整拍照姿态，得到最佳匹配照片，并用该最佳匹配照片得到的分数进行验证。

本发明提出一种带反馈的连续声纹认证方式。这一新方式在认证过程中实时给出实时反馈，从而指导用户用以最佳匹配的发音来进行验证，以减少对真实用户的错误拒绝，提高用户体验。具体而言，开始认证以后，用户可以连续发音，在发音过程中可以实时观察到当前发音的匹配分值，基于此可调整发音方式以提供更好的配合。当得到的最佳匹配得分超过设定的域值时，即可对用户身份进行认定；反之，如果用户在较长时间内（如10秒）无法超过域值，则对其身份进行拒绝。

图2给出这一认证方式的示意图。在认证开始时，用户得分较低，得到这一分数反馈后，用户调整了发音方式，因而得分提高，并在规定时间内通过认证。

图2：带反馈的连续认证方式：认证发起后，用户连续读出屏幕上的滚动内容，（左）认证发起之初，分数较低；（中）用户通过调整发音方式，提高了分数；（右）最优匹配结果通过身份认证。

## 实现样例

本专利目的是保护上述带反馈的连续声纹认证方式。为说明其可行性，我们给出一种基于短时声纹特征向量的实现方案。

前述连续声纹认证方式的核心是计算测试语音中每个短时片段对目标说话人的匹配分值，从而可指导用户即时调整发音方式。我们已经在专利201710214666.0中提出了一种短时声纹特征提取方法，在该方法中，基于深度神经网络，可以在帧级别（0.3秒）提取其声纹特征向量。图3给出语帧上提取声纹特征向量的示意图。



图3：将语音信号转化成声纹特征向量。上图：语音原始信号；下图：声纹特征矩阵。其中v(t)是第t个语音帧所对应的特征向量。

设验证语音第t帧的声纹特征向量为v(t)。注册用户x的注册声纹特征向量为V[x]，则可通过计算v(t) 与V(x)之间的距离给出在t时刻对注册用户x的匹配得分s(t)如下：

s(t) = D(v(t), V(x))

上式中的D(.,.)是两个向量间的距离函数。这一距离函数可取任何形式，只要可以反映两个向量的相似程度即可。可选择的距离函数包括Cosine距离，经过LDA后的Cosine距离，或PLDA距离等。



图4： 基于声纹特征向量的即时分数计算方法。在第t帧，提取该帧语音，基于深度声纹特征提取方法得到声纹特征向量v(t)，与注册用户x的注册发音向量V(x)计算距离D(v(t),V(x))，即可得到在第t时刻的发音匹配分值s(t)，反馈给用户以调整发音方式。

## 发明优点

1. 带反馈的连续认证方式通过给用户提供即时反馈，使真实用户有机会调整发音方式，可显著减小错误拒绝率，提高用户体验。
2. 带反馈的连续认证方式在达到足够置信度时（或者分数过低，或者分数过高）即可做出接受/拒绝的决策，只有对那些置信度不足的个案才延长认证时间。这一方式避免了传统认证系统中所有人在同样长度语音上进行认证的缺陷，平衡了认证精度和认证便宜程度之间的矛盾。