北京大学
PEKING UNIVERSITY

一种基于MFCC和GMM的声纹认证系统实现

北京大学软件与微电子学院信息安全工程2017秋季课程项目开题报告

汇报人 **曹路，李晖，杜思佳，储贤**

1.选题背景

a)根据用户知道的信息进行认证

文本口令、图像口令

容易被肩窥、猜测、字典攻击、重放攻击、木马攻击

随着个人账号增多，记忆成本增加

单因素身份认证无法满足互联网身份认证安全性需求

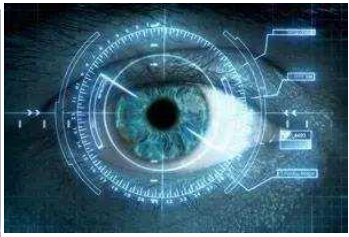
b)根据用户拥有的东西进行认证

动态令牌、智能卡、USB KEY

c)根据用户具有的生物特征进行认证

生物特征：指纹、声纹、掌型、虹膜、脸型等

行为特征：签名、行走步态等



部分生物识别进行身份认证需要高昂的采集设备，比如指纹、虹膜、脸部识别

应用范围限制: 指纹、虹膜、脸部识别等无法用在电话委托系统、电视遥控等应用场景

人在讲话时使用的发声器官在尺寸和形态方面的差异很大具有一定的辨识度

1.选题背景

特征	指纹	掌型	视网膜	虹膜	人脸	静脉	声纹
易用性	高	高	低	中等	中等	中等	高
准确率	高	高	高	高	高	高	高
成本	高	非常高	非常高	非常高	高	非常高	低
用户接受度	中等	中等	中等	中等	中等	中等	高
远程认证	不可	不可	不可	不可	不可	不可	可以
手机采集	部分可	可以	不可	不可	可以	不可	可以

优点:

- 采集方便、自然，用户接受程度较高
- 采集成本低廉（一个麦克风）
- 适用于远程身份确认，通过通讯网络或互联网即可实现远程登录
- 声纹认证的算法复杂度一般较低
- 不易被窃取

缺点：

- 环境噪音对识别存在干扰
- 易受身体状况、年龄、情绪等的影响

总的来说，声纹认证具有良好的特性和广泛的应用场景，优点大于缺点

1.选题背景

公共安全

利用声纹技术进行重点人员监测反电信诈骗或恐怖行为预警等

互联网金融

声纹技术让用户信息真实可靠，预防和遏制互联网金融信息造假

移动支付

声纹ID让用户可在任何时间任何场景下完成支付操作

智能硬件产品

让智能硬件产品“闻声识人”，为用户提供定制化的智能服务

2.产品现状

相关公司：

- SPEAKIN
- 科大讯飞
- 清华灵云
- Google
- Microsoft

Microsoft

iFLY

相关应用：

- 微信声音锁
- Hey, Siri
- 腾讯声纹识别VPR
- 灵云声纹识别VPR

2.产品现状

a) Microsoft Speaker Recognition API

识别较为精准，处理速度快

口令串标准，易被攻击者利用，付费使用，非开源

b) Bob (Idiap Research Institute, Switzerland)

Idia开发的一套声纹识别工具箱

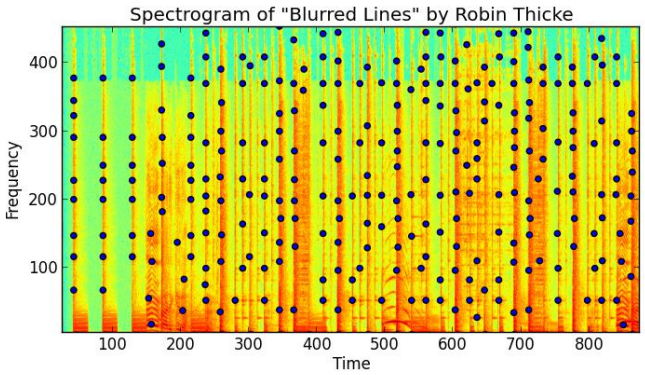
工具箱安装会占用大量空间和内存

语音处理的时长较长

c) Dejavu (Will Drevo, Python and Numpy)

轻量、易用、准确的Python音频指纹库

MIT的Will Drevo基于Python and Numpy开发



3.研究现状

声纹识别核心：特征提取和建模技术

常用的特征提取技术

主要基于短时语音帧分析

- 倒谱分析
- 梅尔频率倒谱分析
- 线性预测（现已不常用）

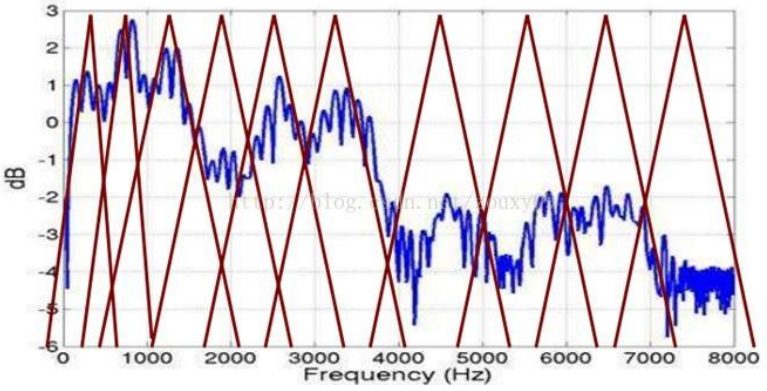
常用的建模方法

- 高斯混合模型（GMM）
- 隐马尔科夫模型（HMM）
- 支持向量机模型（SVM）
- 矢量量化模型（VQ）
- 人工神经元网络（ANN）

HMM常被用做与文本相关的声纹识别

GMM、SVM、VQ主要用做与文本无关的声纹识别

其中GMM模型被认为是现在最优秀的建模方法



随着高斯混合模型技术的快速发展，现在运用最多的技术是使用

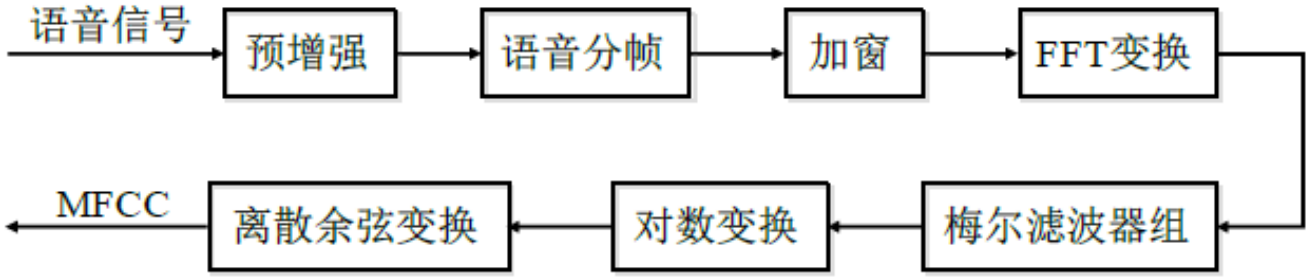
MFCC特征向量+GMM模型

3.研究现状

- ***MFCC：梅尔频率倒谱系数 Mel-Frequency Cepstral Coefficients***

Mel频率是基于人耳听觉特性提出来的，它与Hz频率成非线性对应关系

MFCC则是利用它们之间的这种关系，计算得到的Hz频谱特征，一定程度上模拟了人耳对语音的处理特点



使用Dejavu的python_speech_features实现音频信号的MFCC特征的提取

- ***GMM：高斯混合模型 Gaussian Mixture Model***

GMM模型是单一的高斯密度函数的扩展，可以逼近任意形状的概率密度分布，被广泛应用到语音识别领域

语音信号的特征向量被提取之后，需要对提取的特征向量建模模型训练根据特征参数建立高斯混合模型GMM

使用Python中Sci-kit learn实现的GMM中的fit方法，可得到GMM模型

4.设计思路

用户注册过程

- ✓ 输入用户名和文本口令
- ✓ 记录用户环境噪声，记录时长为5秒
- ✓ 从语料库中随机抽取8-10个单词构成口令串，用户需朗读该单词串进行语音注册（3次）
- ✓ 使用LTSD拟合音频轨迹，消除背景噪音干扰
- ✓ 对去噪后的音频进行语音识别，将识别结果与临时口令串进行模糊匹配，判断阈值，85分通过
- ✓ 提取梅尔频率倒谱系数MFCC特征并正则化
- ✓ 为注册用户建立高斯混合模型GMM

用户登录认证过程

- ✓ 待认证用户输入用户名和文本口令
- ✓ 记录环境噪声，时长为5秒
- ✓ 生成临时口令串，提示待用户朗读口令串
- ✓ 背景噪声去除
- ✓ 对去噪后的音频进行语音识别，将识别结果与临时口令串进行模糊匹配，判断阈值
- ✓ 提取音频信号的MFCC特征、建立GMM模型,与注册用户数据库中的GMM模型进行对数似然估计，进行阈值判断
- ✓ 决定是否授权用户登录

5.安全机制

a) 与文本有关&与文本无关

与文本有关：在用户注册时就确定识别所用的发音内容，由于文本内容是已知的，攻击者则可以通过悄悄录音、诱导用户说指定文字等手段，窃取到用户的登录声纹

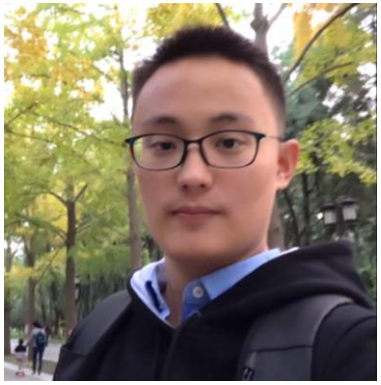
与文本无关：用户在注册、登录时使用的临时口令串均为随机生成，通过提取用户的声音特征与注册数据库中的特征进行匹配，攻击者即使对用户进行录音也无法窃取用户的声纹信息从而登录系统

b) 中间人攻击

攻击者可能监听系统的客户端和服务器端，当用户进行注册时，客户端会将用户音频文件发送回后台服务器端，**中间人在这时会拦截该音频文件，并将其篡改成自己的音频文件发送回服务器端**

解决思路：客户端每次将音频文件发送给服务器端之前，进行数字签名（生成SHA256散列），并将音频文件连同数字签名一起发送给服务器端，若中间人对该音频文件进行篡改，则服务器端会验证失败

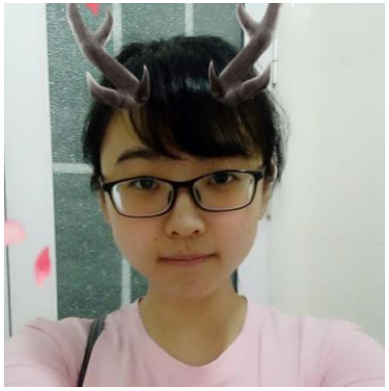
6.项目分工



软件工程与数据技术系

曹路

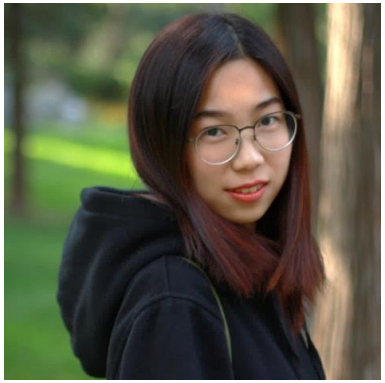
- 课题总体调研
- 认证系统设计
- 开题报告撰写
- 展示PPT撰写
- 系统后续功能实现



网络软件与系统安全系

李晖

- 课题背景调研
- 开题报告撰写
- 展示PPT撰写
- 系统后续功能实现
- 后期系统测试



金融信息与工程管理系

杜思佳

- 声纹认证原理调研
- 开题报告撰写
- 展示PPT撰写
- 系统后续功能实现
- 后期系统测试

网络软件与系统安全系

储贤

- 产品与国内外研究调研
- 开题报告撰写
- 展示PPT撰写
- 系统后续功能实现
- 后期系统测试

7.项目计划

- 2017.10.11 明确项目选题
- 2017.10.18—2017.10.20 查阅相关文献，明确调研方向
- 2017.10.21 小组讨论调研进展，确定下一步内容
- 2017.10.21—2017.10.23 小组讨论，撰写开题报告和展示PPT
- 2017.10.24 提交开题报告和展示PPT
- 2017.10.25 展示PPT
- 2017.10.25以后 基于开题报告，开发相应系统，实现对应的功能
- ……

基于Python，Numpy，*Dejavu*，Sci-kit Learn等

实现一个可用于用户注册或者登录的网页版声纹认证系统

Q&A

感谢聆听！

Thank You For Your Listening

北京大学
PEKING UNIVERSITY