

# 《声纹识别应用》检测报告



总相似率：34.32%

## 基本信息

文档名称	声纹识别应用
报告编号	61c62182-8cd4-4dc3-b92e-448c56e32d2c
文档字数	24613
提交人姓名	张立飞
提交方式	粘贴文本检测
检测范围	学位论文库（含硕博）、学术期刊库、会议论文库、法律法规库、互联网资源库、自建比对库
提交时间	2018-03-20 13:40:00

## 检测报告指标详情

原创率	抄袭率	引用率	字数统计	参考文献字数
65.68%	34.32%	0%	24613	-

## 相似片段位置图



注：红色部分为论文相似部分，蓝色部分为论文引用部分

## 相似片段详情（仅显示前10条）

序号	篇名	来源	相似率
1	声纹识别系统关键技术研究	学位论文库	2.8%
2	基于声纹识别的说话人身份确认方法的研究	学位论文库	1.98%
3	基于MFCC特征和GMM模型的说话人识别系统研究	学位论文库	1.16%
4	VHF无线电话控制单元设计与实现	学位论文库	0.89%
5	复杂野外环境下机器人障碍物检测方法研究	学位论文库	0.88%
6	基于物联网的火电机组远程诊断服务实践	学术期刊库	0.78%
7	What my deep model doesn't know...   Yarin Gal - Blog   ...	互联网资源库	0.71%
8	基于SVM的汉语语音情感识别研究	学位论文库	0.61%
9	... "regedit" 回车打开注册表编辑器,定位到HKEY_LOCAL_MACHI_...	互联网资源库	0.49%
10	映像劫持(Image File Execution Options)的解决方案_百度文库	互联网资源库	0.49%

## 文档原文标注

基于声纹识别的门禁应用

摘要

声纹识别是语音识别中的一种，又被人成为说话人识别。近年来，随着人工智能逐渐进入大众的视线，基于神经网络的身份认证方法得到了进一步的发展。例如人脸识别，虹膜识别以及声纹识别。这些身份认证方法逐渐从理论阶段演化为实际的应用。例如苹果公司在2017年底发布的Face ID – 人脸认证识别功能，其改善了传统密码认证跟你指纹认证的方式，无需记住任何密码组合就能够快速解锁iPhone手机。还有例如微软公司在2015年提出的Windows Hello认证应用；其利用微软公司提供的身份信息采集设备将预登陆者的信息录入到系统中，通过嵌入式的算法进行预登陆者的判断，从而实现快速登陆并开始工作。

身份认证技术随着科技的发展不断得简化认证步骤，优化认证效率，为信息安全提供着保护，为身份认证增加了便捷。

本课题利用高斯混合模型，将从麦克风中采集到的语音信号进行MFCC特征参数提取，并放入模型中训练，计算出预登陆者与语音库中人员的匹配度，从而实现身份认证功能。

关键词：声纹识别 人脸识别 Apple 身份认证 高斯混合模型

?

Windows hello application based on voice print recognition

Abstract

Voiceprint recognition is one kind of the voice recognition, and it is also recognized by the speaker. In recent years, with the artificial intelligence gradually entering the public eye, the authentication method based on the neural network has been further developed. For example, face recognition, iris recognition, and voiceprint recognition. These identity authentication methods gradually evolved from the theoretical stage into practical applications. For example, Apple's Face ID – Face Recognition and Recognition feature released at the end of 2017, which improves the way traditional password authentication and your fingerprint authentication can quickly unlock your iPhone without having to remember any combination of passwords. There is also the Windows Hello authentication application, for example, that Microsoft proposed in 2015; it uses the identity information collection device provided by Microsoft to record the information of the pre-landing person into the system, and uses the embedded algorithm to perform the judgment of the pre-landing person, thus realizing Log in quickly and start working.

With the development of technology, identity authentication technology continues to simplify the authentication process, optimize the authentication efficiency, provide protection for information security, and increase the convenience for identity authentication.

In this project, Gaussian mixture model is used to extract the speech signals collected from the microphone for MFCC diagnosis parameters and put into the model for training. The matching degree between the pre-landing person and the person in the speech library is calculated to realize the identity authentication function.

Key Words : Voiceprint recognition Face recognition Apple identity authentication Gaussian mixture model

?

目录

1. 引言 8

1.1. 声纹识别的研究背景与意义	8
1.2. 声纹识别的发展历史与现状	9
1.3. 声纹识别概述	10
1.3.1. 声纹识别的基本原理和系统结构	10
1.3.2. 声纹识别常用的特诊参数	10
1.3.3. 声纹识别常用的建模方法	11
1.4. Windows Hello概述	12
1.5. 本论文的主要工作	12
2. 开发工具和开发技术	13
2.1. 开发环境	13
2.2. 开发技术	13
2.2.1. 声纹识别技术	13
2.2.2. Socket传输技术	14
2.2.3. 后台刷新	14
3. 系统设计	16
3.1. 系统配置设计	16
3.2. 详细设计	16
3.2.1. Windows下桌面应用系统设计	16
3.2.2. Windows Hello应用设计	16
3.2.3. Web实时监控日志系统设计	16
4. 语音信号预处理	17
4.1. Windows下的语音信号采集	17
4.2. 语音数据提取	17
4.3. 分帧与加窗	19
4.4. 预加重	20
4.5. 端点检测	21

5. 语音信号特诊参数提取	24
5.1. 短时帧能量和短时过零率	24
5.2. Mel频率倒谱系数 ( MFCC )	24
6. 语音信号训练模型	28
6.1. 高斯混合模型(GMM)概述	28
6.2. 单高斯模型	28
6.3. 高斯混合模型	29
7. 声纹识别方法	33
8. 桌面应用	35
8.1. MFC界面设计	35
8.2. 实验数据及结果	36
9. Windows Hello应用	38
9.1. 在Windows登陆界面运行自定义程序	38
9.2. 系统应用流程	38
10. Web实时监控日志系统	40
10.1. 基于本地的日志系统	40
10.2. 基于远程的web日志系统	40
结语	42
参考文献	43
致谢	46
?	
引言	
声纹识别的研究背景与意义	

随着时代的发展，信息技术与网络通信的普及，人类逐步进入大数据的时代。人们生活方式的改变，生活节奏的逐渐加快，使得身份验证的数字化、快捷化显得越来越重要。传统的以密码为基础的身份验证技术存在很大的弊端，很难再满足现在身份认证中安全性和实效性的要求。然而随着信息科学的快速发展，近年来发展起来的生物认证技术正成为一种更加便捷、更加安全，更加隐形化的信息安全技术开始在身份识别邻域展露锋芒。

声纹识别就是在这样一种情况下应运而生。语言是人类所具有的自然属性之一，是人类进行相互通信和交流的最方便快捷的一种手段，每个人说话都具有各自的生物特征，比如，发音器官的先天性生理差异，后天形成的个人发音习惯等行为差异，因此两个人语音是不可能完全相同的；同时，声纹识别对硬件设备的成本要求不高，对声音的采集只需要一台麦克风或者笔记本就可以完成，在对声音的处理阶段，包括采样、量化，特征提取，模式匹配等也只需要在电脑设备上即可完成。所以通过分析说话人语音的方法来对说话人进行识别就成为了可能，这种方法就是声纹识别。

声纹识别相比其他生物特征识别相比，例如指纹识别、人脸识别、虹膜识别等。使用声纹识别具有更加简便、准确、经济及可扩展性良好等各种优势，这些优势体现在：

声纹识别能够在用户不知不觉中就能够做完身份验证，自然得获取说话人的语音，不会涉及到隐私问题，容易被使用者接受。

认证方式简单，只需要用户提供足够的语音信息即可，不用像指纹或虹膜识别技术需要将人体器官贴近信息采集仪器。

实现成本非常低廉，相比于各种指纹或虹膜认证技术需要昂贵的扫描仪器设备，声纹识别只需要简单的语音输入设备就可以了。

在远距离的身份识别中。声纹识别更快捷，只需要通过电话或手机即可实现远距离身份确认。

由于声纹识别具有以上优势，声纹识别的前景可以从以下几个领域概述：

随着计算机的发展，如果将语音作为一种交互手段在某种情况下取代键盘和鼠标，人与机器之间的交流也会越来越自然、流畅。

网络购物已经深入人们的日常生活，网上付款时人们很大程度上依赖于密码或二维码，但是随着各种不同场合的频繁使用，网络安全的隐患越来越明显，如果将声纹识别的概念应用到网络安全中，每次验证系统会随机提供提示文本，能够更有效防止复制和剽窃。

在交通方面，人们在驾驶汽车和飞机时，如果能够使用语音进行系统的控制，取代繁琐的操作，可降低驾驶员劳动强度，减轻驾驶疲劳，对驾驶员的培训系统也会更加简单。

## 声纹识别的发展历史与现状

上个世纪30年代，伴随着信息技术和计算机技术的发展，通过仪器可以实现说话人信息的识别。“声纹”的概念最早由Bell实验室的L.G.Kesta在观察语谱时提出的。人们的研究中心从听音识别和人耳的听辨实验转移到提取有利的声纹特征上来。

上个世纪40年代至70年代是声纹识别技术的创新阶段，Bell实验室的S.Pruzansky提出了基于模式匹配和概率统计方差分析的说话人识别方法，实现了人耳听辨到自动识别技术的越变，各国的专家学者开始研究提出了倒谱技术和线性预测分析技术。

上个世纪70年代至80年代，声纹识别技术的研究重点在于对声音中个性特征参数的非线性或线性处理技术以及寻找新的更有效的模型匹配方法。比如动态时间规划、神经网络、支持向量机、隐马尔可夫模型等方法。

上个世纪90年代至今，各种模式匹配方法逐步成型，高斯混合模型技术作为声纹识别系统较为前沿的方法，收到越来越多的关注，已经成为专家研究的热点。而声纹识别技术也逐渐开始走向人们的视线。

国内研究声纹识别起步较晚，目前取得较好研究成果的机构中有中科院声学所、中科院自动化所

、清华大学、北京大学、中国科学技术大学、上海交通大学等。近年来实验室环境下的声纹识别技术已发展成熟。在实际环境下的声纹识别技术收到越来越多的研究者关注，尤其是在噪声环境下的声纹识别。所以，抗噪是声纹识别的关键技术问题之一。

在相关知识库搜索声纹识别时，早期人们都是使用隐马尔可夫模型来完成模式匹配的工作，而近年来使用高斯混合模型来实现声纹识别的人越来越多，所以，本课题采用MFCC+GMM（Mel倒谱系数+高斯混合模型）的方式来完成。

## 声纹识别概述

### 声纹识别的基本原理和系统结构

声纹识别又称为说话人识别，从目的和判决方式来分类，可以分为两类：说话人确认和说话人辨认。说话人确认用来确定这段话是不是某一个特定的人说的，属于“一对一”问题。说话人辨认用来判断这段话是众多人中其中哪一个人说的，属于“多选一”问题。根据实际情况，不同的工作和任务会采用不同的声纹识别技术。例如网络交易时，需要采用确认（一对一）的技术；刑侦侦察过程中，为了缩小范围，需要采用辨认（一对多）的技术。

声纹识别作为语音识别中的一部分，双方存在很大的差异：语音识别通常关注的是说话的人说了些什么，而不关注说话人的声道变化和说话习惯，更不关注说话的人是谁；声纹识别关注的是说话人是谁，而忽略说话人说什么话，更注重说话人的说话习惯和声道变化等个性特征。声纹识别的基本原理是利用说话人的语音为每一个说话人建立一个能够描述此说话人特征的模型，作为此说话人语音信号特征参数的标准模板，然后和采样过来的语音信号进行对比，以达到判断说话人身份的目的。（如图片1）

### 声纹识别常用的特征参数

在说话人识别系统中，原始语音信号预处理后，需要提取其特征参数，即提取一些适合分类的某些信息特征。在声纹识别样本模型的训练和特定说话人的识别都是基于所选的特征参数进行的。

由于说话内容的语义特征、说话人的情绪波动以及外界环境的干扰，到现在为止还没有找到稳健的特征参数。语音信号的特征参数从各个方面体现了说话人的个性、语言和语义特性。由于在声纹识别系统中，存在语音信号的易变形、训练样本数量的不确定性、外界噪声信号的干扰性等问题，声纹识别系统的特征参数提取方法需要进一步改进和提高。声纹识别的特征参数基本上可以分为三类：

线性预测系数及其派生系数。例如：线性预测系数（LPC）、线性预测倒谱系数（LPCC）、及其组合参数。

由语音频谱直接导出的参数。例如：共振峰、梅尔频率倒谱系数（MFCC）、感知线性预测系数（PLP）。

混合参数。混合参数是由上述不同的特征参数所组成的特征矢量。

不同的特征参数对声纹识别系统的影响不同，决定着系统的性能。一个稳定成熟的特征参数不但可以使声纹识别系统的识别率有所提高，而且可以使声纹识别系统的鲁棒性和稳定性有所增强。

### 声纹识别常用的建模方法

在声纹识别系统中，当特征参数提取后，要用合适的模型来表征这些特征参数，使得模型能够代表语音信号的特征。因此，模型的选择应从语音信号的类型、期望的性能、计算量及存储量等方面考虑。目前主要的方法有：



## 概率统计方法

语音信号在短时间内是相对稳定的，通过对这些稳定的特性如基音、共振峰频率等进行数学分析，从而选用概率密度函数和均值、方差等相关统计量对这些特征进行分类判决。概率统计建立的模型是采用某种概率密度函数的一组参数作为语音模型。概率统计建立的模型考虑了语音信号的统计特性，从各个方面体现出语音信号的统计信息。本课题采用的方法高斯混合模型就属于概率统计方法。

## 动态时间规划方法 ( DTW )

说话人的语音信号不但具有稳定性，而且还具有时变性。将识别模型与参考模型进行时间对比，根据一定的距离测量得出两个模型间的相似程度。

## 矢量量化方法 ( VQ )

该方法是把所有人的特有文本编写成码本，在识别的过程中，根据此码本对测试文本编码，并把判断标准定位量化产生的失真度。

## 支持向量机方法(SVM)

支持向量机方法是一种新的机器学习方法，已广泛应用在指纹识别、人脸识别等模式识别的研究上。SVM是在统计学习理论的基础上建立起来的一种模型，该模型能够较好地解决小样本学习问题。

## 融合方法

融合方法是把以上分类方法与不同参数进行组合，这样可以使系统的性能有所提升。目前常用的有维数简与识别方法组合以及识别方式的结合。

## Windows Hello概述

Windows Hello是一种生物特征授权方式，让用户可以实时访问自己的Windows10设备。

Windows Hello 通过使用你的脸部、虹膜或指纹等生物特征来解锁设备，这种技术比传统密码更加安全。用户就是可以解锁Windows、应用、数据甚至网站和服务的密钥，而不是使用容易被忘记、被破解或随手写下的一串随机排列的字母或数字。相比使用可被共享的密码，Windows Hello可以帮助用户在不适用密码的前提下安全地为应用、网站和网络授权。这样，用户的服务器上就不会存储任何密码，不给黑客可乘之机。

对于初学者来说，Windows Hello 是一个智能身份认证系统。举例来说，如果用户站在Windows 设备面前，用户只需要露一下脸，动动手指，它会自动完成识别与身份验证，同时直接登入你的 Windows 系统。Windows Hello不仅比输入密码更加方便，也更加安全。

## 本论文的主要工作

第一章，介绍了声纹识别的研究背景与意义，阐述了声纹识别的发展历史与现状，介绍了声纹识别的基本原理和本课题的实验流程，常用的特征参数、模式匹配的方法。

第二章，介绍了项目的开发环境跟重要的开发技术。其中，简要介绍了本文中声纹识别的两种方式。

第三章，介绍了系统的部署配置跟各应用的详细设计结构。

第四章，介绍了Windows下如何进行录音，如何将语音文件转化成需要处理的信号数据。还有对语音信号进行参数提取前的准备工作：分帧、加窗、预加重、端点检测。

第五章，介绍了如何提取本文中我们所需要的特征参数；为端点检测所提供的短时过零率与短时帧能量。为建立高斯混合模型所提供的Mel频率倒谱系数。

第六章，介绍了如何实现并建立高斯混合模型，如何对高斯混合模型进行参数的估计，及训练参数的。

第七章，介绍了如何对结合训练数据，进行说话人的判断。

第八章，介绍了如何使用桌面的准备性应用。

第九章，介绍了如何使用实际的验证应用。

第十章，介绍了如何使用web日志监控系统。

？

开发工具和开发技术

开发环境

操作系统:Windows 10 Professional 1709

集成开发环境:Visual Studio 2010 Professionnal

集成开发环境:IntelliJ IDEA 2017.3 Ultimate

JDK版本:Java SE Development Kit 8.0.144

Web服务器:Apache Tomcat Server 8.36

数据库服务器:Mysql 5.6

开发技术

声纹识别技术

本课题使用高斯混合模型建模比较的方法实现声纹识别的关键技术。该系统现有两种方法判定识别结果。

1 x N方式：获取用户输入的语音信号，将语音信号处理后放入每个库存人员的语音高斯混合模型中，计算出匹配值,统计出最大的匹配值，该匹配值所对应的人员即是对应的识别结果人员。

1 x 1方式：先将标准登陆用户的语音放入标准登陆用户的模型中，计算出概率绝对值，将此数值作为标准值，为其创建高门限和低门限。获取用户输入的语音信号，将语音信号处理后放入标准登陆用户的模型中，计算出对象贬值，如果新贬值在高门限与低门限范围内，则确定其为目标用户。

两者的差异在于第一种方式是对象用户与库用户组的比较，表象为一对多，第二种方式为对象用户与标准登陆用户的比较，表象为一对一。适用范围不一样，第一种方式的识别结果为某人，第二



种方式的识别结果为是与否。

## Socket传输技术

Socket又称"套接字"，应用程序通常通过"套接字"向网络发出请求或者应答网络请求。建立网络通信连接至少需要一对端口号(socket)。socket本质是编程接口(API)，对TCP/IP的封装，TCP/IP也要提供可供程序员做网络开发所用的接口，这就是Socket编程接口；

利用Socket可以在不同的机器、不同的平台间可靠地传递消息。本课题利用Socket在Windows客户端中将识别的数据集、识别的结果、识别的流程日志和识别的对象语音发送给服务器。从而实现记录日志，监控系统的功能。

### 图片 2 Socket传输数据集流程

#### 后台刷新

本课题存在的两个部分，Windows客户端与web实时日志监控系统，Windows客户端将识别的结果发送web实时日志监控系统时，日志管理系统后台需要将信息实时的反映在Web前端界面当中。这需要Web客户端浏览器与Web服务器后台实时连接，当Web服务器后台接收到数据时，通知Web客户端浏览器实时刷新。这就需要使用WebSocket让Web客户端与Web服务器实时连接。原理同Socket传输技术。

?

#### 系统设计

#### 系统配置设计

本课题的部署配置如图片3所示，本课题分为三大部分：桌面应用、Window Hello应用、web实时日志系统。

桌面应用：桌面应用用于测试用户设备当前是否存在合适的录音设备，并录下标准用户的语音，计算好声纹信息文件提供给后期识别认证时使用。

Windows Hello应用：Windows Hello应用通过Windows的登陆模式，模拟门禁系统开放与闭合的情况。

Web实时日志系统：Web实时日志系统用于监控当前认证设备的使用情况，防止错误识别情况的发生。

如图3：桌面应用接受用户输入的语音进行计算，生产个人声音信息配置文件，Windows Hello应用获取到标准用户的信息文件与陌生用户的语音生产的信息配置文件数据信息进行对比，从而判断是否该登陆至目标系统。同时，Windows Hello应用与web实时日志系统进行通信，发送认证与匹配信息实时反馈给标准用户。

#### 详细设计

#### Windows下桌面应用系统设计

#### Windows Hello应用设计

#### Web实时监控日志系统设计

?

## 语音信号预处理

### Windows下的语音信号采集

在Windows下可以使用系统自带的API来完成录音操作；使用系统自带的API有两种选择方式，一种是Windows Multimedia API(MMAPI)，这种方式是利用WAV录音固定的格式，先设置好音频测试设备的硬件参数，在开辟一个缓冲区，由录音设备按照设置的参数定次数定位长向缓冲区写入录音数据，再由用户定时读取缓冲区的数据，最后将读取的数据存入文件；还有一种是Media Control Interface (MCI)，使用这种方式，用户可以简单的使用字符串命令实现录音和放音。本课题中采用MMAPI的方法来获取语音文件。

用MMAPI可以实时获得音频数据，MMAPI可以把音频缓冲起来并逐块发送给使用者，这种固定大小的音频裸流数据简称为AudioFrame。当用户预备开始录音时，MMAPI以一种格式打开波形输入设备，发送设备开启消息给回调函数并准备缓冲区，将缓冲区添加到设备，告诉录音设备开始录音。录音期间不停发送数据信息给回调函数。当使用者需要结束录音时，MMAPI告知设备录音结束并发送消息给回调函数处理最后的数据，然后释放缓冲区，最后关闭设备。

### 语音数据提取

每一个语音文件（此处专指wav格式）都必定有四个模块（如图2），它们分别为RIFF块、格式块、附加块、数据块，其中RIFF、格式块和附加块的总长度必定是24或26字节，数据块长度不定。RIFF块有如下内容：标识符、文件长度、WAV标志；格式块有如下内容：标识符、格式块长度、格式类别、声道数、采样频率、数据传送速率、样本字节数、样本位数、附加信息，其中附加信息的有无是通过格式块长度来确定的，如果格式块长度为24则附加信息不存在，如果格式块长度为26则存在附加信息；附加块有如下内容：标识符、块长度、哈希值；数据块有如下内容：标识符、数据长度、数据内容。我们在这里主要要用到数据传送速率、样本字节数、样本位数、数据长度跟数据五个部分的内容。

首先按照文件格式读取文件中我们所需要的内容，读出的数据是以字节的方式存放的，此时的数据并不一定是语音处理的数据，我们需要根据采样位数合并字节，重新换算成语音信号的数据。此处我们实验的数据都是采样位数为16bit的，下面我们就说明以16bit的数据对原字节进行变换的算法：

采样位数为16bit，也就是说其由两个单字节的数据组成，这个两个字节分为高字节数据与低字节数据。判断如果高字节的数据大于等于0转到步骤2如果高字节的数据小于0转到步骤3

判断低字节的数据，如果低字节的数据大于等于0，则新生成的数为高字节的数乘256加上低字节的数，；如果低字节的数小于0，则新生成的数为高字节的数乘256加上那个低字节的数的绝对值加上128，；

判断低字节的数据，如果低字节的数据大于0，则新生成的数为高字节的数的绝对值乘256加上低字节的数的负数，；如果低字节的数小于等于0，则新生成的数为高字节的数的绝对值乘256加上低字节的数的绝对值加上128的负数，；

注：A表示高字节数，B表示低字节数，C表示生成的数

此时的数据就是语音信号可参与转换的数，其数值的范围 此时的数据已经可以参与运算了，但是从实验的角度上，相对过大的数据将会给计算机带来相对较大的负荷，所以此时我们对数值进行“归一化”，将每个数据除以，使得最后的数据范围在-1~1之间。

## 分帧与加窗

语音信号是随时间而变化的，是一个非平稳态过程，但是在短时间范围内（一般在10-30ms内），其特性基本保持不变，即相对稳定，所以语音信号具有短时平稳性。任何语音信号的分析 and 处理必须建立在“短时”的基础上，即进行分帧处理。分帧是通过可移动的有限长度窗函数进行加权的方法实现的。如图三示意图，图中为语音信号分了两个窗，窗一与窗二，窗一的开始与窗二的开始被称为窗移，若将此图放大若干倍，窗与窗之间数值的变化不大，然而却能够体现语音信号的特征。如图四，就是放大后的图。

分帧通常每帧取10~40ms的数据，其中人们常使用25ms作为标准配置，但考虑到后续工作需要，对数据进行快速傅里叶变换，当中需要每帧的数据个数为 $2^n$ ，所以，本课题采用的帧长为256、帧移为125。我们通常使用汉明窗作为加窗函数，汉明窗中的数据能够为傅里叶变换提供便利减少数据量，汉明窗函数： $(2-1)$

其中 $n$ 表示了当前数据在窗中处于第 $n$ 个数据， $N$ 表示汉明窗的窗长度。

## 预加重

研究表明，语音信号的平均功率谱受声门激励与口鼻辐射的影响，高频率约在800Hz以上按照6dB/倍频程跌落，导致语音信号低频段能量大，高频段信号能量明显小，而鉴频器输出噪声的功率谱密度随频率的平方而增加（低频噪声大，高频噪声小），造成信号的低频信噪比很大，而高频信噪比明显不足，从而导致高频传输衰弱，使高频传输困难，因此对语音信号进行预加重，把信号的高频部分进行加重，去除口唇辐射的影响，增加语音的高频分辨率，提高信噪比，改善信号的传输质量。

预加重使用具有6dB/倍频程的提升高频特性的预加重数字滤波器来实现，它一般是一阶的数字滤波器：

$$H(z)=1-\mu z^{-1} \quad (2-2)$$

其中 $\mu$ 为预加重系数，通常取(0.9, 1.0)。经过预加重处理后的结果为：

$$y(n)=x(n)-\mu x(n-1) \quad (2-3)$$

## 端点检测

端点检测是声纹识别中的重要过程，端点检测是从一段语音中判断出语音的起点和终点。声纹识别系统中使用端点检测，一方面可以减少数据量，节约不必要的时间，另一方面可以减少噪声信号的干扰，使声纹识别更准确。再语音识别过程中，尤其是当针对孤立词语识别系统，正确的判定词语的起末端点为模型匹配和提升识别率起着不可替代的作用，既可以将信号处理时间削弱至最小，又可以清理掉无声段的噪声影响，最终提高系统的识别新能，在一定程度上端点检测的正确与否直接影响着整个辨识系统的性能。

双门限端点检测法是音频信号处理当中常用的检测手段，其需要对语音信号计算出时域范围内的两个特征参数（计算方法见第三章：语音信号特征参数提取）：短时过零率和短时帧能量。其结合了短时过零率和短时能量的优点，因此对语音信号的起点和判决更加有效。（实验证明双门限端点检测法只在信噪比较高的情况下具有较高的鲁棒性）双门限法就是在检测的过程中为短时帧能量和短时过零率均设置两个门限，一个高门限和一个低门限，低门限往往对信号的变化反应比较敏感，而当信号达到一定强度才能超过高门限，则可认定是进入语音段。具体做法如下：

计算并得到最大短时能量、最小时能量、最大短时平均过零率和最小时平均过零率就是语音信号的波动范围，下面依据上述的四个范围分别为短时帧能量与短时平均过零率设置双门限（分为

高门限与低门限)，门限的数值将直接影响到最后端点检测判断的结果。在这里，我们先将高门限设置成高范围的1/4，将低门限设置成高范围的1/8。还要设置一个最短语音长度，用于判断是否为爆破音。转到步骤2。

按照窗的个数同时遍历短时帧能量与短时平均过零率，初始状态为静音状态（语音的三种状态，分别为“静音状态”、“过渡状态”、“语音状态”）。（转换图如图5）转到步骤3。

此时语音处于静音状态，如果当前帧的短时帧能量高于所设定的短时帧低门限或者当前帧的短时平均过零率高于所设定的短时帧平均低过零率，则标记当前为语音的起始段转到步骤过渡状态。转到步骤4

此时语音处于过渡状态，如果当前帧的短时帧能量小于所设定的短时帧低门限且当前帧的短时平均过零率小于所设定的短时帧平均低过零率，则状态不变（静音状态）。如果当前帧的短时帧能量高于所设定的短时帧高门限或者当前帧的短时平均过零率高于所设定的短时帧平均高过零率。则标识语音转到步骤语音状态。转到步骤5

此时语音处于语音状态，如果当前帧的短时帧能量小于所设定的短时帧低门限且当前帧的短时平均过零率小于所设定的短时帧平均低过零率且当前的时间长度小于最短时间门限则标识为噪音，否则标识为结束点，然后记录保存下起始点，结束点信息。转到步骤2直至遍历结束。

双门限端点检测流程图如图7.

语音信号特征参数提取

短时帧能量和短时过零率

语音信号的能量不是一直不变的，并且各个音之间的能量存在差异，所以可以判断出语音信号的个性特点。而短时过零率是指每帧内语音信号通过横轴的次数。在有时间横轴上可以看到连续语音信号的波形。若离散语音信号相邻的采样存在不同的代数符号，将此现象成为发生了过零，因此可以计算过零的次数。

用 $E_n$ 代表第 $n$ 帧语音信号 $x(m)$ 的短时能量：

( 3-1 )

用 $Z(n)$ 代表第 $n$ 帧的过零率：

( 3-2 )

其中 $\text{sgn}[]$ 为符号函数，即

$\text{sgn}[x]=\begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$  ( 3-3 )

$w(n)$ 为窗函数（见2.3 分帧与加窗）。

Mel频率倒谱系数（MFCC）

人的听觉系统对声音频率的感知是非线性的，人耳相当于一个滤波器组，它能让人在各种变异环境下正常分辨声音，其滤波作用在对数频率尺度上进行，对1000Hz以下的频率声音的感知呈近似线性关系；而对于1000Hz以上频率声音的感知遵循在对数频率坐标上的近似线性关系，使人耳对低频信号更敏感。

Mel频率滤波器组就是由此实验得到的，其用于模拟人耳对不同频率语音的感知。Mel频率倒谱系数是用一个在低频区域交叉重叠的三角形滤波器组-Mel滤波器组对语音信号的能量谱进行带通滤波。Mel滤波器组在信号的低频区域分布较密，中频区域分布较少，高频区域分布较为稀疏，单个滤波器的带通带宽较大，因此，高频区域的频率分辨率较低，频谱信息较弱，导致信息遗漏。

Mel频率倒谱系数的提取过程如下（见图9）：

将每帧的数据进行快速傅里叶变换（FFT），将数据从时域转换为频域。傅立叶原理表明：任何连续测量的时序或信号，都可以表示为不同频率的正弦波信号的无限叠加。而根据该原理创立的傅立叶变换算法利用直接测量到的原始信号，以累加方式来计算该信号中不同正弦波信号的频率、振幅和相位。有些信号在时域上是很难看出什么特征的，但是如果变换到频域之后，就很容易看出特征了。这就是很多信号分析采用快速傅里叶变换的原因。另外，快速傅里叶变换可以将一个信号的频谱提取出来，这在频谱分析方面也是经常用的。在此处就是利用快速傅里叶变换将信号的频谱提取出来。

$$X_K = \sum_{n=0}^{N-1} [X_n e^{-i2\pi kn/N}] \quad \text{离散傅里叶变换 (DFT)} \quad (3-4)$$

$$X_n = 1/N \sum_{k=0}^{N-1} [X_k e^{i2\pi kn/N}] \quad \text{逆向离散傅里叶变换 (IDFT)} \quad (3-5)$$

求频谱的平方，得到谱线的能量：

$$E(i, k) = [x(i, k)]^2 \quad (3-6)$$

其中*i*表示第*i*帧，*k*表示第*k*条谱线

将谱线能量通过Mel滤波器组

求最大/最小Mel频率，及滤波器的中心间距

$$\text{melFremax} = 1125 \log(1 + \text{SampleRate}/(2 \times 700)) \quad (3-7)$$

$$\text{melFremin} = 1125 \log(1 + 0/700) \quad (3-8)$$

$$\text{?}_\text{mel} = (\text{melFremax} - \text{melFremin}) / (\text{filterNum} + 1) \quad (3-9)$$

其中melFremax表示的是最大Mel频率，melFremin表示的是最小Mel频率，SampleRate表示的是信号的采样频率，filterNum表示的是帧的个数。

求mel滤波器组中每个滤波器的实际频率位

$$f[i] = \text{floor}(((N+1) \times 700 \times (\exp((\text{melFremin} + \text{?}_\text{mel} \times i)/1125) - 1)) / \text{SimpleRate}) \quad (3-10)$$

*I*为当前滤波器 melFremin为最小Mel频率

*N*为帧长 ?<sub>mel</sub>为滤波器的中心间距

SimpleRate为采样频率

将谱线能量通过滤波器

$$S(i, m) = \sum_{k=0}^{N-1} E(i, k) H_m(k), 0 \leq m < M \quad (3-11)$$



$$H_m(k) = \begin{cases} \frac{(k-f(m-1))/(f(m)-f(m-1)) \& f(m-1) \leq k \leq f(m)}{1} & \text{if } (f(m+1)-k)/(f(m+1)-f(m)) \& f(m) \leq k \leq f(m+1) \\ 0 & \text{otherwise} \end{cases} \quad (3-12)$$

I表示第i帧 m表示第m个滤波器

N表示帧长 k表示帧内第k个数据

M表示滤波器的总个数

求对数

$$\text{mel}(i, m) = \ln(\text{mel}(i, m)) \quad (3-13)$$

进行离散余弦变换

$$C(i, n) = \sum_{m=0}^{M-1} S(i, m) \cos[\pi n(m+0.5)/M] \quad 0 \leq n \leq L \text{ 且 } 0 \leq m < M \quad (3-14)$$

其中i为帧数，n为阶数 表示第i帧中的第n阶

M为第m个滤波器 M为总滤波器的个数

S(i, m)为[i][m]位置上的滤波能量

?

语音信号训练模型

高斯混合模型(GMM)概述

在声纹识别系统中，模型的建立至关重要，不同的模型建立方法对识别的性能影响不同。在进行声纹识别时，将输入系统的识别语音特征参数和已有的说话人识别模型进行相似比对分析，然后根据对比分析结果对待识别的说话人身份做出相应的判断。因此，声纹识别模型的建立与特征参数提取同等重要，也是声纹识别的关键之一。传统的建模方法：如动态时间规划、人工神经网络等，这些建模方法计算量较大，训练样本的时间较长，缺乏统计特性。高斯混合模型能较好地描述样本空间的稳步，能够对任意形状的分布进行近似。

近年来调查发现，说话人的特征分布并没有严格遵循某一个特定分布，例如高斯分布等；但是几乎所有的分布都能够使用高斯分布的混合加权值来近似接近，从而获得了高斯混合模型。基于高斯混合模型的声纹识别技术的基本理论为针对训练话者集合内的每一个说话人构建属于自己身份特征的概率分布模型，这种概率模型中所设计的参数值是由说话人自身的特征参数分布情况所决定，所以可以用以描述说话人的身份特性。为了方便处理，假设所有话者概率密度函数具有相同的形式，但函数中参数值设定不同，此时说话人模型可以看成为符合给定概率密度分布形式的一组参数值。

高斯混合模型是单一高斯概率密度函数的延伸，GMM能够平滑地近似任意形状的密度分布。高斯混合模型种类有单高斯模型（SGM）和高斯混合模型（GMM）两类。类似于聚类，根据高斯概率密度函数（PDF）参数不同，每一个高斯模型可以看作一种类别，输入一个样本x，即可通过PDF计算其值，然后通过一个阈值来判断该样本是否属于高斯模型。很明显，SGM适合于仅有两类别问题的划分，而GMM由于具有多个模型，划分更为精细，适用于多类别的划分，可以应用于复杂对象建模。

单高斯模型



单高斯模型的多维高斯（正态）分布概率密度函数定义如下：

$$N(x; \mu, \Sigma) = 1/\sqrt{((2\pi)^d |\Sigma|)} \exp[-1/2 (x-\mu)^T \Sigma^{-1} (x-\mu)] \quad (4-1)$$

$x$ 是维数为 $d$ 的样本向量（列向量）， $\mu$ 是模型期望， $\Sigma$ 是模型方差。

假设训练样本属于类别 $C$ ，则上式变为：

$$N(x/C) = 1/\sqrt{((2\pi)^d |\Sigma|)} \exp[-1/2 (x-\mu)^T \Sigma^{-1} (x-\mu)] \quad (4-2)$$

其表示了样本属于 $C$ 的概率大小。若将任意测试样本 $x_i$ 输入式（4-2），均可以得到一个标量 $N(x_i; \mu, \Sigma)$ 。然后根据贬值 $t$ 来确定该样本是否属于该类别。（贬值 $t$ 的确定：可以为经验值，也可以由实验确定。另外可以先另 $t=0.7$ ，以0.05为步长一直减到0.1左右，选择使样本变化最小的哪个贬值作为最终 $t$ 值）

单高斯分布模型的几何意义：单高斯分布模型在二维空间近似于椭圆，在三维空间上近似于椭球，因为在很多场合下，同一类别的样本点并不满足“椭圆”分布的特性。所以才引入了高斯混合模型。

## 高斯混合模型

高斯混合模型是单一高斯机率密度函数的延伸，由于 GMM 能够平滑地近似任意形状的密度分布，因此近年来常被用在语音、图像识别等方面，得到不错的效果。

例如有一批观察数据 $X = \{x_1, x_2, \dots, x_n\}$ ，数据个数为 $n$ ，在 $d$ 维空间中的分布不是椭球状，那么就不适合以一个单一的高斯密度函数来描述这些数据点的机率密度函数。此时我们采用一个变通方案，假设每个点均由一个单高斯分布生成（具体参数未知），而这一批数据共由 $M$ （明确）个单高斯模型生成，具体某个数据属于哪个单高斯模型未知，且每个单高斯模型在混合模型中占的比例未知，将所有来自不同分布的数据点混在一起，该分布称为高斯混合分布。

从数学上讲，我们认为这些数据的概率密度分布函数可以通过加权函数表示：

$$P(x_i) = \sum_{j=1}^M [a_j N(x_i; \mu_j, \Sigma_j)] \quad \sum_{j=1}^M a_j = 1 \quad (4-3)$$

$$N(x; \mu_j, \Sigma_j) = 1/\sqrt{((2\pi)^d |\Sigma_j|)} \exp[-1/2 (x-\mu_j)^T [\Sigma_j]^{-1} (x-\mu_j)] \quad (4-4)$$

$a_j$ 是权值因子，其中的任意一个单高斯分布 $N(x; \mu_j, \Sigma_j)$ 叫做这个模型的一个component，只要 $j$ 取得足够大，这个xx Mixture Model就会变得足够复杂，就可以逼近任意连续的概率密度分布。

高斯混合模型是一种聚类算法，每个component就是一个聚类中心。即在只有样本点，不知道样本分类的情况下，计算出模型参数 $(x; \mu, \Sigma)$ 。其计算方法就是EM算法。

用训练好的模型去差别样本所属分类，分为两个步骤：

随机选择 $k$ 个component中的一个（被选中的概率是 $a_j$ ）。

把样本带入刚选好的component中，判断是否属于这个类别，如果不属于则返回第一步。

现在假设我们由 $N$ 个数据点，并假设他们服从某个分布，现在需要确定里面的一些参数的值。例如：在GMM中，我们就需要确定 $(x; \mu, \Sigma)$ 这些参数，我们的想法是：找到这样的一组参数，它所确定的概率分布生成这些给定的数据点的概率最大，而这个概率的乘积则称为似然函数。

$\prod_{i=1}^N P_r(x_i; \theta)$  似然函数 (4-5)

通常单个点的概率很小，连乘之后数据会更小，容易造成浮点数下溢，所以一般取对数，公式(4-5)变成：

$\sum_{i=1}^N \log(P_r(x_i; \theta))$  称为log-likelihood function (4-6)

高斯混合模型的log-likelihood function为：

$\sum_{i=1}^N \log[\sum_{k=1}^K a_k N(x_i; \mu_k, \Sigma_k)]$  (4-7)

要找到最值的模型参数，使GMM的log-likelihood function的期望最大，使用E-M算法求解。

EM算法的基本思路是：随机初始化一组参数 $\theta$ ，根据后验概率 $P_r(Y|x; \theta)$ 来更新Y的期望 $E(Y)$ ，然后用 $E(Y)$ 代替Y求出新的模型参数 $\theta$ ，如此迭代直到 $\theta$ 趋于稳定。

**E-Step**：假设模型参数已知的情况下隐含变量Z分别取 $Z_1, Z_2, \dots$ 的期望，亦即Z分别取 $Z_1, Z_2, \dots$ 的概率。在GMM中求数据点由各个component生成的概率。

$r(i, k) = b_k P_r(Z_k | x_i; a, \mu, \Sigma)$  (4-8)

权值因子 $b_k$ 表示在训练集中数据点属于类别 $Z_k$ 的频率，在GMM中它就是 $a_k$ 。

$r(i, k) = (a_k N(x_i | \mu_k, \Sigma_k)) / (\sum_{j=1}^K a_j N(x_i | \mu_j, \Sigma_j))$  (4-9)

**M-Step**：用最大似然的方法求出模型参数，认为 $r(i, k)$ 就是数据点 $x_i$ 由component k生成的概率。

$N_k = \sum_{j=1}^N r(i, k)$  (4-10)

$\mu_k = 1/N_k \sum_{j=1}^N r(i, k) x_i$  (4-11)

$\Sigma_k = 1/N_k \sum_{j=1}^N r(i, k) (x_i - \mu_k)(x_i - \mu_k)^T$  (4-12)

$a_k = N_k / N$  (4-13)

?

## 声纹识别方法

对于一个声纹识别系统，假设整个系统共有N个说话人，其对应的M阶的高斯混合模型参数分别为： $\gamma_1, \gamma_2, \dots, \gamma_N$ 。在辨识阶段，给定一个待识别的语音样本的特征矢量序列 $X = \{x_1, x_2, \dots, x_T\}$ ，分别定义如下几个函数： $P(\gamma_n)$ ， $P(x)$ ， $P(x|\gamma_k)$ ， $P(\gamma_k|x)$ 。

$P(\gamma_n)$ 为第n个人的先验概率

$P(x)$ 为所有人的特征矢量集的概率密度

$P(x|\gamma_k)$ 为第n个人的特征矢量集的概率密度

$P(\gamma_k|x)$ 为后验概率，表示在特征矢量集为x的条件下，确定说话人为第n个人的概率。

则这段语音属于第n个说话人的最大后验概率为：

$$P(\gamma_n | x) = P(x | \gamma_n) P(\gamma_n) / P(x) = P(x | \gamma_n) P(\gamma_n) / (\sum_{m=1}^N P(x | \gamma_m) P(\gamma_m)) \quad (5-1)$$

识别结果 $n^*$ 可以由最大后验概率给出，即

$$n^* = \arg \max_{1 \leq n \leq N} [P(\gamma_n | x)] \quad (5-2)$$

假定该语音信号出自封闭集合，则每个人的先验概率相同，则：

$$P(\gamma_n) = 1/N \quad 1 \leq n \leq N \quad (5-3)$$

即识别结果等价于：

$$n^* = \arg \max_{1 \leq n \leq N} [P(x | \gamma_n)] \quad (5-4)$$

为了使计算更加简便，通常采用对数似然函数

$$L(x | \gamma_n) = \ln P(x | \gamma_n) \quad (5-5)$$

得到判决准则为：

$$n^* = \arg \max_{1 \leq n \leq N} \sum_{t=1}^T \ln P(x_t | \gamma_n) \quad (5-6)$$

其表示在N个人中能使 $\sum_{t=1}^T \ln P(x_t | \gamma_n)$ 最大的第n个人。

其中：

$$P(x) = \sum_{i=1}^M \left[ \beta_i \frac{1}{\sqrt{((2\pi)^S |\Sigma_i|)}} \exp[-1/2 (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)] \right] \quad (5-7)$$

$\beta_i$ 为加权系数 N为库存人数

$\mu_i$ 为均值向量 M为SGM的个数

$\Sigma_i$ 为协方差矩阵 S为GMM的维数

?

桌面应用

MFC界面设计

本课题使用MFC进行界面设计，对录音API与高斯混合模型训练识别进行整合设计完成了一个完整的带有界面的声纹识别系统。下面介绍一些MFC常用的标准控件。

表格 1 使用的Windows标准控件

控件 MFC类 描述

按钮 Cbutton 用来展示某种行为的按钮，以及复选框

编辑框 Cedit 用于键入文本

列表 ClistCtrl 显示文本及其图标列表的窗口

组合框 CcomboBox 编辑框与列表的组合

静态文本 Cstatic 常用语为其他控件提供标签

界面设计如图13所示。

通过录音模块中的录音按钮，可以即时录音，软件会选择操作系统默认的录音设备开始录音，录音格式wav文件（16bit，16000hz，单声道）。录音前可在“录音人”的编辑框中填写录音人名。  
如果在录音完毕后录音人编辑框中有文字信息，则将录音人信息写入进文件。

在语音文件与模型文件中用户能够看到库中的录音跟已经训练完成的数据。

用户可以随机选择语音文件中的某一项进行训练，即将录音人的高斯混合模型参数写入文件存入库中。

用户可以随机选择语音文件中的某一项进行识别，即将待识别人与语音库中的人进行人员识别。

#### 实验数据及结果

实验样本数为5，每个样本中说话人所说的内容皆为“我们需要帮助”。训练每个人的语音到语音库中，每次选择其中个语音样本进行识别，看能否确认这个语音样本是否属于他本人。

样本wangzhe 文件大小42144B 帧数165

Wangzhe Zhanglifei Zhaozouxian Liuchang Zhaoquan

$P(x|)$  -713.795 -1056.2 -1320.7 -1451.75 -2773.97

样本zhanglifei 文件大小96044B 帧数：380

Wangzhe Zhanglifei Zhaozouxian Liuchang Zhaoquan

$P(x|)$  -3326.97 -2166.06 -4949.54 -3154.88 -8358.6

样本zhaozouxian 文件大小43574B 帧数171

Wangzhe Zhanglifei Zhaozouxian Liuchang Zhaoquan

$P(x|)$  -1113.54 -3443.29 -724.993 -5008.76 -3609.02

样本liuchang 文件大小47652B 帧数181

Wangzhe Zhanglifei Zhaozouxian Liuchang Zhaoquan

$P(x|)$  -1080.08 -1085.65 -1275.73 -640.919 -3005.28

样本zhaoquan 文件大小45630B 帧数177

Wangzhe Zhanglifei Zhaozouxiang Liuchang Zhaoquan

P(x|) -1069.84 -1250.03 -1249.39 -949.363 -2898.08

在5个样本中样本1、2、3、4都成功识别，只有样本5识别结果为同是女生的Liuchang。整个系统识别率达到80%。

?

## Windows Hello应用

Windows Hello是操作系统Windows 10中新提出的一种生物验证方法。它可以通过采集登陆者的数据进行后台验证，从而跳过输入密码的过程。

在本课题中，我们采用麦克风来采集登陆者信息，如果验证成功，则将密码组合成键盘消息发送给系统登陆程序（logonUI.exe）；如果验证失败，则弹出提示信息。

在Windows登陆界面运行自定义程序

在系统中运行“regedit”打开注册表，定位到注册表的  
HKEY\_LOCAL\_MACHINE\SOFTWARE\Microsoft\WindowsNT\CurrentVersion\Image File  
Execution Options；

在[Image File Execution Options]上点击右键，选择[新建]-[项]，将其命名为 utilman.exe；

点击[utilman.exe]，在右侧空白处单击右键，选择[新建]-[字符串值]，命名为[Debugger]；

双击打开Debugger，将其数值数据框中输入要打开程序的路径，如[任务管理器  
]：C:\Windows\System32\taskmgr.exe；

设置完成后，在锁定界面点击“轻松使用”将不会打开菜单，而是启动修改后的程序；从而在Windows登陆界面运行声纹识别程序。

## 系统应用流程

在系统进入登陆界面时能够启动该应用。应用启动后会立即启动系统默认的录音设备开始采集说话人的信息。接着应用程序会休眠5500ms，在此段时间之内，录音线程会自动将采集到的信息写入指定的缓冲区内。当休眠结束，应用调用停止录音操作，录音线程将所有采集到的信息写入文件中供后面使用。（录音文件长度约5s，其中500ms用于启动并设置录音设备用。）

获取到录音文件后，应用开始读取文件数据，将数据提取出来进行信号初始化并计算信号的特征参数。然后将库存中的数据提取出来，创建高斯混合模型，将预识别人的特征参数放入库存中每个人的高斯混合模型中，计算最大概率。如果最大概率是预先设置的设备使用者的数据集，那么系统会向LogonUI.exe发送登陆密码的组合消息。否则，则弹出提示。从而实现快速登陆。

?

## Web实时监控日志系统

为了监控系统的状态，捕获异常性的错误信息，web实时监控日志系统记录了所以程序运行时的状态并上传至服务器。该web实时监控系统分为两个部分：基于本地的日志系统。基于远程的日志系统。

## 基于本地的日志系统

当Windows Hello应用程序执行时，会自动向程序存放的位置中新建名为“logs”的文件夹（如果文件夹不存在），并以当前日期的形式新建文本文件。在程序的执行过程中，关键模块的衔接部分或异常性的信息会标明时间同时写入至文件当中。当用户需要时可以随时打开目录文件夹查看应用的执行过程跟执行情况。

## 基于远程的web日志系统

本地日志系统只适用于用户能够访问验证机器的情况。但是远程web日志系统能够让已经获得访问许可的人查看日志信息。如果将验证机器通过内网穿透映射至公网内，即使在千里之外也能够查看信息。

将web日志系统发布到web服务器上，通过连接地址打开网站输入登陆账户跟密码即可访问到日志系统。用户可以查看跟删除日志信息，从而实现远程化。

?

## 结语

本文研究的内容是基于MFCC倒谱系数和高斯混合模型GMM的说话人识别系统。其中主要是对说话人识别所采用的特征向量进行了研究，给出了MFCC倒谱系数的具体提取过程和算法。还有高斯混合模型的研究，最终将高斯混合模型与实际应用相结合起来，完成了这样一个Windows Hello的声纹识别系统。

起初想到这个题目是从朋友那里听来的“语音签名”这个词汇。原本是准备制作医疗信息系统（因为原先的指导老师做了医疗信息系统公司的高级技术顾问，准备利用现有条件开发产值的高新技术系统），但我个人的意思是将“语音签名”的本质声纹识别结合身边的应用做出一个发展。所以我就想到了Windows Hello这样一个概念。

经过一个学期的准备与三个月的制作，浏览了44篇论文，参考了68篇网络技术博客，终于完成了本课题。这个系统最终只是一个实验系统，期待未来能够真正的与微软公司合作，加入声纹识别这样一个认证方式。

最后，希望有一天，说话人识别技术能够真正成熟起来，与其他技术结合起来在现实生活中大放异彩。

?

## 参考文献

基础框架部分：

[1] 杨胜跃，周宴宇. 语音信号端点检测方法与展望. [J].通信技术, 2005，卷号（07）：0005-04 .

[2] 刘波.基于短时能量和过零率分析的语音端点检测方法研究[R].湖北武汉：武汉理工大学信息工程学院，2004 .

[3] 李爱平，党幼云. VQ声纹识别算法和实验. [J].西安工程科技学院学报, 2007，卷号（21）：0848-04 .



[4] 周跃海, 童峰. 采用DTW算法和语音增强的嵌入式声纹识别系统. [J].厦门大学学报, 2012, 卷号 ( 51 ) : 0174-05 .

[5] 于娴, 贺松. 基于GMM模型的声纹识别模式匹配研究. [J].通信技术, 2015, 卷号 ( 48 ) : 0097-05 .

[6] 王正创. 基于MFCC的声纹识别系统研究[D]. 无锡: 江南大学, 2014 .

[7] 鲁晓倩, 关胜晓. 基于VQ和GMM的实时声纹识别研究. [J]., 2014, 卷号 ( 23 )

[8] 赵峰, 于洋. 基于VQ和HMM 的双层声纹识别算法. [J].桂林电子科技大学学报, 2017, 卷号 ( 37 ) : 0008-07 .

[9] 臧晓笠. 基于基于高斯混合模型GMM的说话人识别方法. [J].科技信息, 2006

[10] 辽宁工业大学.基于高斯混合模型的声纹识别方法及系统[P].CN : 102324232 A , 2011 .

[11] 张超琼, 苗夺谦. 基于高斯混合模型的语音性别识别. [J].计算机应用, 2007, 卷号 ( 28 ) : 0360-03 .

[12] 茅剑, 林奇. 基于声纹识别的嵌入式防盗系统. [J].计算机与现代化, 2009, 卷号 ( 11 ) : 0163-03 .

[13] 周雷. 基于MFCC的声纹识别的说话人身份确认方法的研究[D]. 上海: 上海师范大学, 2016 .

[14] 杨阳, 陈永明. 声纹识别及其应用. [J].语音技术, 2007, 卷号 ( 02 ) : 0045-02 .

[15] 裴鑫. 声纹识别系统关键技术研究[D]. 哈尔滨: 哈尔滨理工大学, 2016 .

[16] 朱浩冰, 郭东辉. 声纹识别系统原理及其关键技术. [J].网络安全, 2007, 卷号 ( 09 ) .

[17] 古今, 郭立. 一种基于感知特征的鲁棒性语音认证算法. [J].中国科学院研究生院学报, 2009, 卷号 ( 26 ) : 474-482 .

[18] 古今. 语音感知认证的关键技术研究[D]. 合肥: 中国科学技术大学, 2009 .

特征参数部分:

[19] 周萍, 李晓盼. 混合MFCC特征参数应用于语音情感识别. [J].计算机测量与控制, 2013, 卷号 ( 21 ) : 1996-03 .

[20] 郭春霞, 裴雪红. 基于MFCC的说话人识别系统. [J].电子科技, 2005, 卷号 ( 11 ) : 11-012 .

[21] 韩一, 王国胤. 基于MFCC的语音情感识别. [J].重庆邮电大学学报, 2008, 卷号 ( 20 ) : 0597-06 .

[22] 丁爱明. 基于MFCC和GMM的说话人识别系统研究[D]. 南京: 河海大学, 2006 .

[23] 王恩泽, 何东健. 基于MFCC和双重GMM的鸟类识别方法. [J].计算机工程与设计, 2014, 卷号 ( 35 ) : 1868-04 .

[24] 王萌, 王福龙. 基于端点检测和高斯滤波器组的MFCC说话人识别. [J].计算机系统应用, 2016, 卷号 ( 25 )

[25] 何朝霞, 潘平. 说话人识别中改进的MFCC参数提取方法. [J].计算机技术与工程, 2011, 卷号 ( 11 )

训练模型部分：

[26] 向昌盛. 高斯混合模型心音信号自动识别. [J].南京理工大学学报, 2016, 卷号 ( 40 ) : 0560-06 .

[27] 蔡桂林 . 高斯混合模型用于语音情感识别研究[D] . 广西：广西师范大学，2016 .

[28] 翟玉杰 . 基于GMM-SVM说话人识别的信道算法研究[D] . 吉林：吉林大学，2015 .

[29] 陈银燕 . 基于HMM和GMM天然地震与人工爆破识别算法研究[D] . 广西：广西师范大学，2011 .

[30] 张奇, 苏鸿根. 基于高斯混合模型的乐器识别方法. [J].计算机工程, 2004, 卷号 ( 30 ) : 0133-02 .

[31] 杨澄宇, 赵文. 基于高斯混合模型的说话人确认系统. [J].计算机应用, 2001, 卷号 ( 21 ) : 0007-02 .

[32] 余清清, 李应. 基于高斯混合模型的自然环境声音的识别. [J].计算机工程和应用, 2011, 卷号 ( 25 ) : 152-155 .

[33] 任民宏, 鲁秋菊. 基于高斯混合模型自适应肤色识别算法. [J].陕西理工学院学报, 2016, 卷号 ( 32 ) : 0053-04 .

[34] 许百林 . 基于矢量量化 ( VQ ) 和高斯混合模型 ( GMM ) 的说话人识别的研究[D] . 南京：东南大学，2005 .

[35] 刘恒, 吴迪. 运用高斯混合模型识别动物声音情绪. [J].应用天地, 2016, 卷号 ( 35 ) : 520-20 .

性别识别部分：

[36] 肖汉光, 何为. 基于MFCC和SVM的说话人性别识别. [J].重庆大学学报, 2009, 卷号 ( 32 ) : 0770-05 .

[37] 旁程, 李晓飞. 基于MFCC与基频特征贡献度识别说话人性别. [J].华中科技大学学报, 2013, 卷号 ( 41 ) : 0108-04 .

[38] 高原 . 基于性别分类的说话人识别研究[D] . 徐州：江苏师范大学，2012 .

？

致谢

请允许我致谢，为这四年一路走来的成长，为我经历和得到的所有。

首先我要感谢带我走进语音信号处理的刘永俊老师，虽然刘老师从来没有对我做出实质性的指导，虽然刘老师常常把我们的成果占为己有，虽然刘老师慢慢变成一个商人而不是一个老师。但我还是得感谢他；我依然记得大二的时候参与的第一个实验室项目，刘老师找了一个划水的项目主持人混了一个校级项目，骗了学校2000块钱。正因为这样，我才能出色得表演我的产品经理天赋。也正因为这样，我才在语音信号中约扎越深。最终在零指导下，单靠查阅资料完成了省级项目。正因为这样，我才知道，原来我竟然有这么强的学习能力。

我还要感谢我的同窗秦立浩，我做语音，他做图像。他其实比我更加适合做研究性的内容，我适合做一个项目经理。虽然他常常幻想担任我的角色，但其实我更喜欢他这样的人才。毕业了，他考研去了，我即将进入公司，我不清楚我的未来是不是还能遇见这样的朋友，但如果再见，我一定让他担任我们公司的研发组组长。

感谢我的指导老师，虽然我在写这篇致谢的时候我才大三，但我知道，未来的你一定是一个认真负责的老师。一定是你，也只有你才能知道我的论文到这么完美的地步。

感谢14软件单招班的所有同学，虽然你们都爱玩游戏，但我相信，未来一定有属于你们的出路。请不要畏惧，向着你们所向往的目标去飞吧，你们是我这24年来的骄傲与辉煌的见证者。

感谢所有教导过我们的老师们，在我心里，你们各有特色，感谢你们的教导，感谢你们！