# 基于微信小程序的校园微贷平台

张泽川

（新疆大学 信息科学与工程学院 830046）

**摘 要**：本系统基于树莓派平台通过计算机语言、麦克风收集语音实现软硬件结合，实现说话人识别功能。说话人识别也称声纹识别,其目的是根据说话人的声音特征来完成说话人的辨认，随着网络信息化技术的迅猛发展,身份验证的数字化、隐性化、便捷化显得越来越重要,说话人识别作为一种生物认证技术，说话人识别技术研究的关键是语音信号的特征提取和模式匹配等问题。本系统主要研究了当前比较常用的语音增强算法、语音特征种类以及声纹识别算法。本系统的构建为：对含噪语音实现语音增强、语音预处理、语音特征的提取、声纹特征库的建立、声纹识别四个模块。实现各个模块的功能并且连接成一个完整的语音系统。

**关键词**：校园微贷平台；微信小程序；

## Speech system based on Raspberry Pi platform

*ZhangZechuan，YangGuang，ShiXiongwei，FangYuan*

(Xinjiang University College of Information Science and Engineering 830046)

**Abstract**：The system is based on the Raspberry Pi platform, collects voice through computer language and microphone, realizes the combination of software and hardware, and realizes speaker recognition function. Speaker recognition is also called voiceprint recognition, and its purpose is to complete speaker recognition based on the characteristics of the speaker's voice. With the rapid development of network information technology, the digitalization, covertness and convenience of identity verification have become more and more Important, speaker recognition is a kind of biometric authentication technology. The key to speaker recognition technology research is the feature extraction and pattern of speech signals. Matching and other issues. This system mainly studies the commonly used speech enhancement algorithms, speech feature types and voiceprint recognition algorithms. The construction of this system includes four modules: realizing speech enhancement, speech preprocessing, speech feature extraction, voiceprint feature database, and voiceprint recognition for noisy speech. Realize the function of each module and connect into a complete voice system.

**Key words**：speech processing; voiceprint recognition; speech enhancement; Gaussian mixture model; Mel cepstrum coefficient; spectral subtraction; raspberry pi

**引 言**

近年来，随着人民生活水平和消费能力的不断提高，

（好处）。

**1 基于微信小程序的平台**

（整体构架）。

**1.1 小程序端**

语音信号分析是语音处理系统的前提和基础，有分析出可表示语音信号本质特征的参数后，才有可能利用这些参数进行高效的语音通信、语音特征提取和语音识别等处理。

贯穿于语音分析全过程的是“短时分析技术”。因为语音信号从整体来看其特性及表征其本质特征的参数均是随时间而变化的，所以它是非平稳态过程，不能用处理平稳信号的数字信号处理技术对其进行分析处理。但是，由于不同的语音是由人的口腔肌肉运动构成声道某种形状而产生的响应，而这种口腔肌肉运动相对于语音频率来说是非常

**1.1.1页面**

（商品或信息的发布页面）涉及到多张图片的上传，异步处理

用户的登录，获取openid，之后与服务器间进行通信。

时间的显示，修改时间数据的格式。

this的指代

①在频域内，语音信号的频谱分量主要集中在300~3400Hz的范围内。利用这个特点，可以用一个防混频的带通滤波器将此范围内的语音信号频率分量取出，然后按8kHz的采样率对语音信号进行采样，就可以得到离散的语音信号。

②在时域内，语音信号具有“短时性”的特点即在总体上，语音信号的特征是随着时间而变化的，但在一段较短的时间间隔内，语音信号保持平稳。在浊音段表现出周期信号的特征在清音段表现出随机噪声的特征。

**1.1.2业务处理函数**

语音信号的预处理包括预滤波、采样和量化加窗、预加重、端点检测等过程。所选用的实验语音数据，利用树莓派的麦克风录制输入，采用16kHz样频率、16bit量化、单声道的PCM录音格式。由于语音信号在帧长为0ms~30ms之内是相对平稳的，同时为了便于计算FFT，本系统选取帧长N为256个语音点，帧移M为128点。



**1.2 服务器端**

语音增强是解决噪声污染的一种有效方法,它的一个主要目标是从带噪语音中提取出尽可能纯净的原始语音,尽可能多的去掉语音信号中的噪声和干扰,以改善它的质量。在进行说话人识别时,语音增强的目的是在信号的时域中对含噪语音进行消噪处理,在保持语音较小失真的同时,有效抑制噪声,提高输出信噪比,有效地提高噪声环境下说话人识别系统识别性能。

**1.2.1谱减法概述**

谱减法在假定加性噪声与语音相互独立的前提下，在频域将噪声的功率谱从带噪语音信号功率谱中减去，得到纯净语音的谱幅度估计，并借助于带噪语音的相位经反傅里叶变换恢复出语音信号。

带噪语音信号，经FFT变换后，有由此可得

由于和独立，所以和独立，而为零均值的高斯分布，所以有

对于一个分析帧内的短时平稳随机过程，有

其中为无语音时的统计平均值，由此得原始语音估计值为

其中是增强后的语音信号的幅度。

令，后验信噪比，则式（1）可改写为

当小于1时，无意义。因此可改写成

是大于零的常数，其物理意义是：对于带噪语音的每一个频谱分量乘以一个系数，信噪比高时，含有语音的可能性大，衰减系数小。反之，则认为含有语音的可能性小，衰减系数大。介于两者之间时根据后验信噪比的大小调整增益大小。

谱减法的优点是比较简单，实时实现比较容易。不足是信噪比范围较窄，在信噪比较低时对语音的可懂度损伤较大，还存在一定的“音乐噪声”。

**1.2.2维纳滤波法概述**

维纳(Wiener) 滤波是适合用来解决这样一类从噪声中提取信号问题的过滤方法，其优点是适应面较广，无论平稳随机过程是连续的还是离散的，是标量的还是向量的都可应用，还可求出滤波器传递函数的显式解，并进而采用由简单的物理元件组成的网络构成。维纳滤波以均方误差最小为准则来解决线性滤波和预测问题的，假设 Wiener 的输入为含噪声的随机信号，期望输出与实际输出之间的差值为误差，对该误差求均方即为均方误差，那么均方误差越小噪声滤除效果就越好。

滤波器的冲激响应是实现均方误差最小的关键，能够满足 Yule-wa1ker 方程就可使维纳滤波器达到最佳。

**1.3 数据库**

对于本项目我们选择使用Mysq数据库，进行存储。

经过MFCC特征参数提取后，各帧语音信号就形成了一个个特征矢量。识别时，将待测语音与模板库中的每一个模板进行模式匹配找到距离最小的模板作为输出结果。

**1.3.1 梅尔倒谱系数概述**

人耳对于音强的解析能力，随着音强的大小而改变，对小的声音具有良好的解析度，所以我们常用dB值来表示音强，也就是对数的方式表示音强，让音强较小的部分放大，音强较大的部分放大，音强较大的部分缩小。人耳在每个频域点上的敏感度都不相同且有临界频带现象，对1KHz以下的音频部分有较高的频率分辨度，临界频带基本成线性关系，宽度大概为100Hz。而1kHz以上的临界频带成指数关系。因此，我们可以根据人耳的这种听觉特性，用梅尔倒谱系数来划分频带，并将一个频带的频率成分看作是能量强度，然后将这些频带强度，将其进行离散余弦变换后就变成倒频谱，这就是梅尔频率倒频谱。MFCC特征参数提取算法流程如图所示：

图2 MFCC特征提取框图

**1.3.2 预加重概述**

MFCC特征提取的第一步就是语音信号高频分量的能量，如果把一帧语音的频谱看作是元音，低频分量的能量比高频分量的多很多。我们开口讲话的时候，在声带发元音时，在声带端看来，是信号通过了一个声带成形滤波器，输出的空气流量速度波形大致有的高频衰减。人耳听到的声音相当于经过了一个低通滤波器，做了 的高频衰减，如果将语音信号通过一个高通滤波器，即可以补偿上述的高频衰减。达到预加重的目的。计算公式如下：

**1.3.3 语音分帧及加窗概述**

对于一段离散的音频信号，加上一个固定窗长度的窗函数得到一帧语音，并对其逐帧处理。因为语音信息具有短时平稳性，窗函数的长度一般取 10~40 ms。为了让帧与帧之间具有连续性，还需要采用分帧的方式，让前后帧有重叠部分。帧移大概是帧长的一半，即 5~20 ms。如果语音信号的采用率是 16kHz，窗的长度取 256 个点，那么一帧的时间是 16 ms，前后帧重叠一半的信号，帧移则是 128 个点，时间为 8 ms。矩形窗和汉明窗是比较常用的两种窗函数。矩形窗的数学公式如下：

语音信号加上矩形窗，窗外的信号被设置为 0，窗内保留原语音信号，优点是计算简单，缺点是造成了语音信号的不连续性。这种不连续性在听觉上会产生瞬时的噪声，从频域上看，语音信号的频谱会被破坏。汉明窗弥补了矩形窗的缺点，窗的两边是缓慢减小，在界限上不会形成显著的间断现象。汉明窗的公式如下：

最后将预加重后的语音信号乘以窗函数，计算公式如下

**1.3.4 DFT傅里叶变换**

将加窗后的一帧语音信号经过 DFT 变换计算得到在频域中每个频率采样点上的值。FFT 变换是 DFT 变换的一种快速运算形式，帧的长度必须为 2 的整数次方，若不足 2 的整数次方，采用补零的方式补足，才可以用 FFT 变换。傅里叶分析是基于欧拉公式：

离散傅里叶变换的公式如下：

**1.3.5 梅尔滤波器组和对数能量概述**

人的听力并不是对全频带都是敏感的，尤其对于大于 1000 Hz 更加不敏感。需要将 FFT 变换后的频谱转换至梅尔刻度的梅尔频域上。梅尔频域和线性赫兹频域之间的映射在 1000 Hz 以下是线性的，1000 Hz 以上是对数性的。梅尔频域上和线性频域之间的转换公式如下：

离散余弦变换（DCT）: 对全部个滤波器输出的对数能量做离散余弦变换（Discrete Cosine Transform），求出 L 阶的梅尔倒频谱系数，这里 L 通常取 12。离散余弦转换公式如下：

式中 为信号的梅尔频率倒谱系数；为三角滤波器的数目。

**1.3.6 多阶差分梅尔倒频谱系数概述**

梅尔倒谱特征向量仅仅是描述了单帧语音信息的功率频谱包络，由于语音信号帧与帧之间是变化的，需要对梅尔倒谱系数进行一阶或二阶差分运算，一阶差分运算表示在相应的梅尔倒谱特征中语音信号帧与帧之间的动态信息，二阶差分运算则表示在对应的一阶倒频谱特征中语音信号帧与帧之间的动态信息。梅尔倒谱系数的一阶差分运算的计算公式如下：

式中为一阶差分梅尔倒谱系数；为梅尔倒谱系数。

二阶差分梅尔倒谱系数计算方式跟一阶的一样，只是将更改为即可。

**1.4高斯混合模型**

高斯混合模型（）利用多个单变量高斯分布的加权和来表征不同说话人的语音特征向量分布状态并以此来区分不同说话人的模型。

由个单高斯分布的加权和构成，表示如下式：

其中为高斯混合模型的参数，为高斯混合模型的阶数，为维的语音特征向量，为每个单高斯变量的权重且。高斯混合模型如下图所示：

图3 高斯混合模型原理图

**1.4.1 模型概述**

一个阶的混合高斯模型本质上是一种多维概率分布函数其概率密度函数是由个单高斯概率密度函数加权和得到的。

是子分布，每个子分布是维的高斯联合概率分布，可表示如下：

其中为个高斯分布，为均值向量，为

协方差矩阵。

由上述内容可以看出，一个任意的分布就可以通过模型的个均值、方差和权重参数来拟合。其中均值参数表示了每个高斯分布的位置，方差表示了每个高斯分布的变化范围,而权重参数则表示每个高斯分布的幅度大小即分布在高斯的数据多少。因此一个可以由下列参数描述：

⑴：模型中高斯密度函数的个数

⑵每个高斯函数的参数，包含均值（维），协方差矩阵（维矩阵）

⑶每个高斯密度函数的权重

这样可以简记一个为，那么构建一个模型也就是对已知样本数据进行训练，给定后估计出参数，， 即完成训练。

**1.4.2 模型参数估计算法**

高斯混合模型中常用到的参数估计的方法是最大似然估计法，在给定特征样本集后最大似然估计的目标就是找到使得高斯混合模型的似然函数最大的模型参数。假定训练样本集为，则的似然函数为：

上式和参数集之间存在很复杂的非线性函数关系,按照常规的求极值的方法来寻求极大值点很困难。一般是通过算法求出最大值。其基本思想是从一个初始模型开始，估计一个新的模型参数,使其新的似然函数值大于原参数值计算的似然函数值，满足下式：

再以为当前参数值，进行下一次迭代，制止满足收敛条件。

算法包括两个步骤：

第一个步骤为，在和已知情况下，计算样本集的对数似然函数的数学期望，定义为下式：

其中为当前参数的估计，用于计算期望值。为估计出的期望值。

第二个步骤为，最大化期望值。计算公式如下：

算法能够保证似然值收敛于似然函数的局部最大点。

**1.4.3模型参数初始化**

具体迭代计算中就涉及到模型参数的初始化取值问题本系统采用K-均值聚类方法将语音数据划分得到个聚类使类内平方和最小。

采用均值法对算法进行初始化,可以提高算法的迭代速度。而K-均值聚类方法也是一种用局部搜索方法来迭代得到最优解。

**1.4.4 说话人识别**

训练后的高斯混合模型中，其中每个说话人都对应一个模型，彼此之间就是参数的不同，记为。在识别阶段,设待识别语音的观测特征向量序列为，则该说话人是训练集中第个人的后验概率为下式：

式中为第个说话人的先验概率，为所有说话人条件下特征向量集的概率，为第个说话人产生特征向量集的条件概率，最终识别结果即为最大的后验概率对应于的说话人。计算如下式：

**2 基于Eclipse软件的服务器开发**

树莓派与常用的单片机相比，不但可以完成IO引脚控制，由于树莓派的运行有相应的操作系统，还可以完成复杂的任务管理与调度，能够支持更上层应用的开发。开发语言的选择不仅仅只限于C语言。树莓派可以连接底层硬件与上层应用，可以实现物联网的云控制和云管理等功能，也可以忽略树莓派的IO控制，使用树莓派搭建小型的网络服务器，做一些小型的测试开发。

同时树莓派与通用的PC平台相比，树莓派可以提供IO引脚控制其他底层硬件的功能，体积和成本都比较小。这种特性更适合我们本科生阶段电子信息工程专业做一些项目的实现和开发。此外还有关于语音的导入、处理等技术。本项目就是针对这几项关键技术，开展我们的学习。这些模块各自独立，却又紧密联系，而且这些模块有着比较完备的理论基础和类似的实践先例，有着比较好的实践性。

**3 数据库设计**

本系统主要基于Python语言。运行环境依赖于树莓派中Linux下的Python环境。系统代码的流程如下图所示：

 

图4 代码流程框图

其中语音降噪采取谱减法；语音提取MFCC作为特征，其提取的MFCC阶数为2。左侧为预先收集语音进行GMM模型的训练，结果生成一个模型参数文件；右侧为对模型的应用，收集语音进行与训练集同样的处理后计算后验概率给出识别结果。

**4 校园平台的使用及改进**

整个系统流程分为两大部分，分别为模型建立和使用模型来进行说话人识别。首先收集语音经过增强及预处理后提取并建立语音特征库，由特征库数据进行模型的训练，保持模型并且进行说话人识别应用。

**4.1谱减法语音增强结果**

图5 输入语

图7 谱减法对加噪语音去噪效果

**4.2****维纳滤波法对信号增强结果**

使用MATLAB平台实现了对生成语音信号加噪并使用维纳滤波

**5 总结**

本项目完成了基于GMM模型的语音说话人识别功能，但存在的问题是：对于文件导入时，由于不同的文件压缩格式不同可能导致函数不能直接运行，因此需要进行处理格式转换。

量机相结合两种办法来提高识别率。

**参考文献:**

1. 冯志民.防录音假冒身份的声纹确认方法[D].上海师范大学,2019.
2. 房安栋.复杂背景下声纹特征提取与识别[D].中南林业科技大学,2014.
3. 王萌.说话人识别算法研究[D].广东工业大学,2017.
4. 于树本.基于MFCC的说话人语音识别系统的研究[J].黑龙江科技信息,2015(27):69-70.
5. 陈强.基于GMM的说话人识别系统研究与实现[D].武汉理工大学,2010.
6. 曾伟浩.数字助听器听觉场景识别和自动增益控制算法研究及实现[D].哈尔滨工业大学,2015.
7. 陈黎.基于SVM和GMM的说话人辨识方法研究[D].武汉理工大学,2011.
8. 陈伯胜.基于VQ和GMM的与文本无关的说话人识别研究[D].重庆大学,2007.
9. 郭万鹏.基于深度学习的说话人识别技术研究[D].兰州理工大学,2019.
10. 容强，肖汉.基于MMSE维纳滤波语音增强方法研究与Matlab实现[J].计算机应用与软件,2015,32(1):153-156.
11. 孟欣.改进的参数自适应的维纳滤波语音增强算法[D].太原理工大学.
12. 宋丽亚.基于小波变换的说话人语音特征参数研究[D].西安电子科技大学,2004.
13. 闵姝君,田岚.一种自适应变阶谱减降噪算法在电子耳蜗中的应用[J].声学学报(3):82-88.