**基于时域分析技术的语音识别**

任泽华 自动化71 学号:2171411498

姬文虎 自动化71 学号:2176413411

程敏敏 自动化71 学号:2171421548

刘雨芊 自动化74 学号:2173724059

秦祯 自动化75 学号:2175021232

2019.10.7

**基于时域分析技术的语音识别问题**

**摘要**

语音是一种典型的、易于获取的一维时序信号。语音信号处理及识别技术是数字信号处理课程的重要应用板块。时间序列分析、快速傅里叶变换、滤波器

设计等多项数字信号处理的教学内容在语音识别核心技术中均占有重要地位。

本次实验面向孤立字语音识别基本任务，由时域角度进行探讨、分析、解决。 本文将通过能识别数字0～9的语音识别系统的实现过程详细阐述基于KNN等算法的特定孤立词识别的相关原理和关键技术.

在信号采集部分，首先通过编程方式运用audiorecorder函数进行了语音信号的采集并建立了班级语音信号库。在预处理部分，我们对语音信号进行了批量处理：运用汉明窗函数进行分帧加窗处理，短时时域特性分析、基于双门限法的端点检测。基于时域分析提取的特征向量，本文选择了KNN、决策树等分类器进行了孤立字语音识别。最后进行了实验对比及量化分析。

运用KNN，最后的识别准确度为50%左右。

【关键词】语音识别 时域分析法 MATLAB 分类器

**一．问题分析**

**1.1问题背景**

自动语音识别技术起源于 20 世纪50 年代，最早的商用系统是 IBM 在 90 年代推出的 ViaVoice。经过半个多世纪的发展，语音识别技术目前已日趋成熟并成功应用到人们的日常生活之中，如苹果手机的 Siri 体验、科大讯飞的迅速崛起等。根据识别的对象不同，语音识别任务大体可分为3类，即孤立词识别（isolated word recognition），关键词识别（keyword spotting）和连续语音识别。其中，孤立词识别 的任务是识别事先已知的孤立的词，如“开机”、“关机”等。本题即属于孤立词识别。

语音是一种典型的、易于获取的一维时序信号，语音信号处理及识别技术也是数字信号处理课程绝佳的实践途径。时间序列分析、快速傅里叶变换、滤波器设计等多项数字信号处理的教学内容在语音识别核心技术中均占有重要地位。本次实验即面向语音识别基本任务，需熟悉语音数据的基本形式及特点，理解并应用离散时间信号的基本分析、处理方法，理解语音识别技术的概貌。

**1.2问题提出**

1. 采集“0”-“9”语音信号
2. 进行语音信号预处理：语音信号的分帧与加窗、语音信号短时时域分析、基于双门限法的端点检测等
3. 编程实现基于时域分析技术实现孤立字语音识别

**二．问题解决**

**2.1 采集“0”-“9”语音信号**

本文在 MATLAB 环境中使用 audiorecorder 函数录制语音信号，采集“0”、“1”…“9”这 10 个语音的 wav 文件，每个类别应采集 10 组以上的样本。为获得更大的样本量，本班建立了班级语音库。

**2.2 语音信号预处理**

**2.2.1 语音信号的分帧与加窗**

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，所以它是一个非平稳态过程，不能用处理平稳信号的数字处理技术对其进行处理。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缓慢的，所以从另一个方面看，虽然语音信号具有时变特性，但是在一个短时间范围内（一般认为在 10~30 ms 的时间内），其特性基本保持不变即相对稳定，因而可以将其看做是一个准稳态过程，即语音信号具有短时平稳性。所以任何语音信号的分析和处理必须建立在“短时”的基础上，即进行“短时分析”，将语音信号分为一段一段来分析其特征参数，其中每一段称为一“帧”，帧长一般取 10-30 ms。这样，对于整体的语音信号来讲，分析出的是由每一帧特征参数组成的特征参数时间序列。 分帧是用可移动的有限长度窗口进行加权的方法来实现的，这就是用一定的窗函数 w(n)来乘s(n)，从而形成加窗语音信号 。窗函数w(n) 的选择（形状和长度），对于短时分析参数的特性影响很大。为此应选择合适的窗口，使其短时参数很好地反映语音信号的特性变化。在语音信号数字处理中常用的窗函数有三种：

本次实验，我们选取汉明窗作为加窗函数。

**2.2.2 语音信号短时时域分析**

语音信号的时域分析就是分析和提取语音信号的时域参数。时域分析通常用

于最基本的参数分析及应用，如语音的分割、预处理、分类等。语音信号的时域参数有多种，本实验着重掌握短时能量、短时平均幅度及短时过零率。

[1]、 短时能量与短时平均幅度

设第n帧语音信号 的短时能量用表示，则其计算公式如下：

是一个度量语音信号幅度值变化的函数，但它有一个缺陷，即它对高电平 非常敏感（因为它计算时用的是信号的平方）。为此，可采用另一个度量语音信号幅度值变化的函数，即短时平均幅度函数 ，它的定义为：

也是一帧语音信号能量大小的表征，它与 的区别在于计算时小取样值和大取样值不会因取平方而造成较大差异，在某些应用领域中会带来一些好处。

[2]、 短时过零率

短时过零率表示一帧语音中语音信号波形穿过横轴（零电平）的次数。对于

连续语音信号，过零即意味着时域波形通过时间轴；而对于离散信号，如果相邻

的取样值改变符号则称为过零。因此，过零率就是样本改变符号的次数。过零率

实质上是信号频谱分布在时域的一种最简单的体现，即高频分量丰富的信号其过

零率也一般较高。

设第n帧语音信号 的短时过零率用 表示，则其计算公式如下：

   

式中，sgn[.]是符号函数

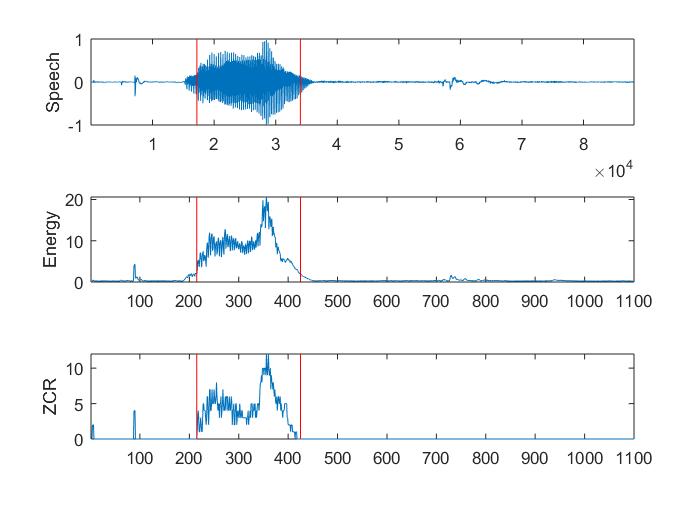
 

其具体计算步骤如下：

a）首先对信号进行去直流化；

b）然后按照时间顺序统计采样点数值符号变号的次数；

c）将上述计数出的次数针对序列时长进行归一化操作，即得到过零率。



图“0”的短时时域分析结果

**2.2.3 基于双门限法的端点检测**

根据语音的统计特性，可以把语音段分为清音、浊音以及静音（包括背景噪

声）三种。语音端点检测本质上是根据语音和噪声的相同参数所表现出的不同特

征来进行区分。在双门限法中，短时能量可以较好地区分出浊音和静音。对于清

音，由于其能量较小，在短时能量检测中会因为低于能量门限而被误判为静音；

短时过零率则可以从语音中区分出静音和清音。将这两种检测结合起来，就可以

检测出语音段（清音和浊音）及静音段。在基于短时能量和过零率的双门限端点

检测算法中首先为短时能量和过零率分别确定两个门限，一个为较低的门限，对

信号的变化比较敏感，另一个是较高的门限。当低门限被超过时，很有可能是由

于很小的噪声所引起的，未必是语音的开始，到高门限被超过并且在接下来的时

间段内一直超过低门限时，则意味着语音信号的开始。

**2.3基于时域分析技术实现孤立字语音识别**

针对提取完成的语音特征向量，选取合适的分类器算法来实现自动语音判别。

选择的分类器包括 Navïe Bayesian、Fisher 线性判别、决策树、支撑向量机、最近邻分类器、KNN等。具体实验情况如下：

* + 1. **选取KNN作为分类器**

KNN是通过测量不同特征值之间的距离进行分类。它的思路是：如果一个样本在特征空间中的k个最相似(即特征空间中最邻近)的样本中的大多数属于某一个类别，则该样本也属于这个类别，其中K通常是不大于20的整数。KNN算法中，所选择的邻居都是已经正确分类的对象。该方法在定类决策上只依据最邻近的一个或者几个样本的类别来决定待分样本所属的类别。KNN算法的结果很大程度取决于K的选择。在KNN中，通过计算对象间距离来作为各个对象之间的非相似性指标，避免了对象之间的匹配问题，在这里距离一般使用欧氏距离或曼哈顿距离：

同时，KNN通过依据k个对象中占优的类别进行决策，而不是单一的对象类别决策。这两点就是KNN算法的优势。

KNN的算法描述为：

（1）计算测试数据与各个训练数据之间的距离；

（2）按照距离的递增关系进行排序；

（3）选取距离最小的K个点；

（4）确定前K个点所在类别的出现频率；

（5）返回前K个点中出现频率最高的类别作为测试数据的预测分类。

**2.3.2 选取支持向量机作为分类器**

**三．实验对比与量化分析**

**参考文献**

[1]

[2]

**附录**