

基于 MATLAB 的语音信号处理教学演示平台的设计

王志鹏，张 雷

(南阳师范学院 物理与电子工程学院,河南 南阳 473061)

摘 要:针对“语音信号处理”课程的特点,将 MATLAB 语言引入课程,介绍了一款语音信号处理教学演示平台.该平台能够很好地帮助学生理解和掌握课程的基本原理和基本分析方法.

关键词:语音信号处理;MATLAB;时域分析;频域分析;线性预测分析

中图分类号:TN 912;G 642.423 文献标识码:A 文章编号:1671-6132(2009)12-0081-04

0 引言

“语音信号处理”是本科电子类专业课程,是研究用数字信号处理技术对语音信号进行处理的一门学科.“语音信号处理”课程内容较多,涉及不少数学运算和公式推导,要求掌握《信号与系统》、《数字信号处理》等课程理论,并且语音信号处理的很多概念学生都是第一次接触到,难以理解.如果仅采用课堂讲授的教学方式学生难以接受,无法调动学生的学习主动性和积极性,教学效果差.所以说,淡化理论的推导过程,使理论结果可视化,概念直观化,将是帮助学生理解教学内容,提高教学效果的关键.

MATLAB 是 MathWorks 公司推出的一套高性能的数值计算和可视化软件,它集数值分析、矩阵计算、信号运算和图形显示于一体,构成了一个方便的、界面友好的用户环境,已被广泛用于信号与图形处理、控制系统设计、通信系统仿真等领域.

针对课程特点,在教学实践中利用 MATLAB 平台,将教学内容中一些比较重要的内容或者难以理解的概念开发成可视化程序.该演示平台包括窗函数分析模块、时域分析模块、频域分析模块、同态分析模块、线性预测分析模块等,基本涵盖了语音信号处理课程的所有内容.在课堂教学过程中使用该演示平台,可以把复杂的数学过程用图形显示出来,使学生对课程的知识有一个更加直观的认识,便于理解.本文就对该教学演示平台进行简要的分析介绍.

1 窗函数

贯穿于语音信号分析与应用全过程的是“短时分析技术”,也就需要利用窗函数对语音信号进行分帧处理,将语音信号分成一帧一帧来分析其特性参数.在语音信号处理中常用的窗函数是矩形窗和哈明窗,它们的表达式分别如下:

矩形窗:

$$w(n) = \begin{cases} 1, & 0 \leq n \leq (N-1) \\ 0, & n = \text{else} \end{cases}$$

哈明窗:

$$w(n) = \begin{cases} 0.54 - 0.46\cos[2\pi n/(N-1)], & 0 \leq n \leq (N-1) \\ 0, & n = \text{else} \end{cases}$$

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,使其短时参数能更好地反映语音信号的特性变化.这里就可以通过教学演示平台生成这两种窗函数的时域波形以及频谱图,如图 1 所示.

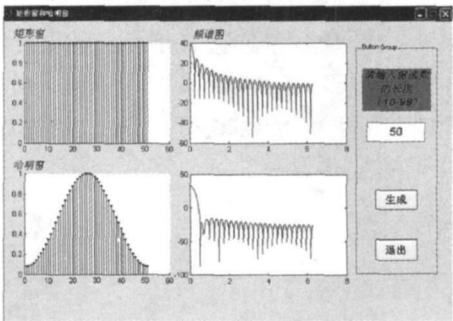


图 1 矩形窗和哈明窗

收稿日期:2009-09-25
作者简介:王志鹏(1982-),河南南阳人,助教,硕士,主要从事无线电导航方面的研究.

从两种窗函数的频谱图中可以清晰地看到, 哈明窗的主瓣宽度比矩形窗大一倍, 即带宽增加一倍, 同时其带外衰减也比矩形窗大得多. 因此可以得到, 矩形窗的谱平滑性能较好, 但损失了高频成分, 使波形细节丢失. 而哈明窗就相反. 同时, 还可以设置窗函数的长度 N 的大小, 来反复观察窗函数长度对于其频谱的影响.

通过上面的演示, 可以很好地展示出窗函数的形状和长度对于窗函数特性的影响, 使学生更好地理解为什么说合适地选择窗函数对于后续语音信号的处理十分重要.

2 语音信号的时域分析

语音信号处理的各种应用的前提和基础是对语音信号的分析. 根据分析的参数的不同, 语音信号分析可分为时域、频域、倒谱域等. 在以上的分析方法中, 最简单、物理意义最明确的就是时域分析. 典型的时域分析包括短时平均能量、短时平均过零数和短时自相关函数等^[1].

语音信号的短时能量分析是基于语音信号能量随时间有相当大的变化, 特别是清音段的能量比浊音段的小得多. 短时平均能量定义 E_n 如下:

$$E_n = \frac{1}{N} \sum_{m=n-N+1}^n [x(m)w(n-m)]^2 = \frac{1}{N} \sum_{m=n-N+1}^n [x(m)w(n-m)]^2.$$

其中, $x(m)$ 表示语音信号; $w(m)$ 是窗函数, 可以是前面所介绍的矩形窗或哈明窗; N 是窗函数的长度.

短时平均能量也即对窗函数所取的其中一帧语音信号幅值取平方再求和, 它的主要用途有:

区分清音段和浊音段; 区分声母和韵母的分界, 有声和无声的分界; 作为超音段信息, 用于语音识别中.

语音信号的短时过零分析是语音时域分析中最简单的一种. 对于连续语音信号, 可以考察其时域波形通过时间轴的情况. 而对于离散时间信号, 如果相邻的取样值改变符号则称为过零. 短时平均过零数 Z_n 的定义为:

$$Z_n = \frac{1}{N} \sum_{m=n-N+1}^n |\operatorname{sgn}[x(m)] - \operatorname{sgn}[x(m-1)]| w(n-m) = \frac{1}{N} \sum_{m=n-N+1}^n |\operatorname{sgn}[x(n)] - \operatorname{sgn}[x(n-1)]| * w(n),$$

式中, $\operatorname{sgn}[\cdot]$ 是符号函数.

短时平均过零数也可以用来区分清音和浊音. 这是由于发浊音时, 尽管声道有若干个共振峰, 但由于声门波引起了谱的高频衰落, 所以其语音能量约集中于 3kHz 以下. 而发清音时, 多数能量出现在

较高频率上. 既然高频率意味着高的平均过零数, 低频率意味着低的平均过零数, 那么可以认为浊音时具有较低的平均过零数, 而清音时具有较高的平均过零数. 利用短时平均过零数还可以从背景噪声中找出语音信号, 可用于判断寂静无声段和有声段的起点和终点位置. 相关分析是一种常用的时域波形分析方法, 它有自相关和互相关的不同, 分别由自相关函数和互相关函数来定义. 相关函数用于测定两个信号在时域内的相似性, 如利用互相关函数, 可测定两个信号间的时间滞后或从杂音中检测信号. 而自相关函数用于研究信号本身, 如信号波形的同步性、周期性等. 这里主要讨论自相关函数. 短时自相关函数 $R_n(k)$ 的定义为:

$$R_n(k) = T[x_n(m)] =$$

$$\sum_{m=-\infty}^{\infty} x(m)w(n-m)x(m+k)w(n-m-k).$$

利用短时自相关函数可以对浊音信号估计它的基音周期. 利用自相关函数的性质, 短时自相关函数可以改写为下式:

$$R_n(k) = \sum_{m=0}^{N-1-k} [x(n+m)w(m)] \cdot [x(n+m+k)w(m+k)],$$

式中, $w(m) = w(-m)$, N 为窗函数的长度.

从上式可以看到, 在传统的自相关函数计算中, 是两个等长的序列进行乘积的和, 这样随着延迟 k 的增加, 进行乘积和的项数在减少, 所以总体上自相关函数的幅度值随着延迟 k 的增加而下降. 因此, 一方面在利用传统自相关函数计算波形周期时, 如果窗长不够长, 包含的周期数不够多, 则会给周期计算带来困难. 另一方面, N 值也要尽可能地小, 因为语音信号的特性是变化的, 如果 N 过大将影响短时性. 为解决这个问题, 可用修正的短时自相关函数来代替短时自相关函数. 修正的短时自相关函数定义为:

$$\hat{R}_n = \sum_{m=0}^{N-1} [x(n+m)w_1(m)] [x(n+m+k)w_2(m+k)],$$

$$w_1(m) = \begin{cases} \text{非零}, & 0 \leq m \leq N-1 \\ 0, & \text{其余 } m \text{ 值} \end{cases}$$

$$w_2(m) = \begin{cases} \text{非零}, & 0 \leq m \leq N-1+K \\ 0, & \text{其余 } m \text{ 值} \end{cases}$$

修正的短时自相关函数是用两个长度不同的窗口, 截取两个不等长的序列进行乘积和, 两个窗口的长度相差最大的延迟点数 K 这样就能始终保持乘积和的项数不变, 即始终为短窗口的长度^[2].

语音信号时域分析的教学演示平台如图 2 所示.

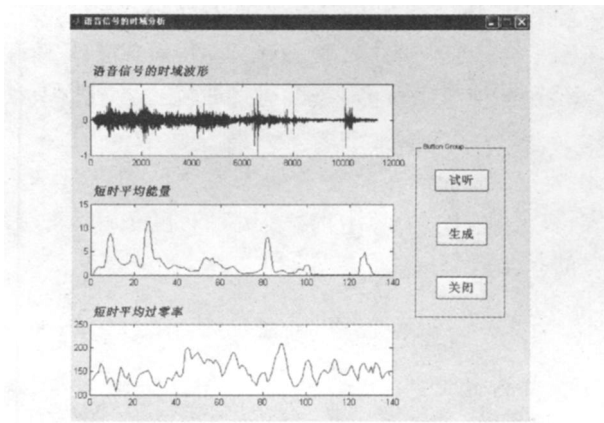


图 2 语音信号的时域分析

3 语音信号的频域分析

傅立叶变换在信号处理中具有十分重要的作用,然而,语音信号是一个非平稳过程,因此,适用于周期、瞬变或平稳随机信号的标准傅立叶变换不能用来直接表示语音信号.所以说语音信号的频域分析方法就是短时傅立叶变换,相应的频谱称为“短时谱”.可以对某一帧语音进行傅立叶变换,即短时傅立叶变换^[3],其定义为:

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-jm\omega}.$$

由定义知,短时傅立叶变换就是窗选信号的标准傅立叶变换.短时傅立叶变换有两个自变量: n 和 ω ; 所以它既是关于时间 n 的离散函数,又是关于角频率 ω 的连续函数.

在数字信号处理中,为了便于计算机分析离散信号的频谱,提出了离散傅立叶变换的概念,与之类似,如令 $\omega = 2\pi k/N$,可以得到离散的短时傅立叶变换,它实际上是 $X_n(e^{j\omega})$ 在频域的取样.其定义如下:

$$X_n(e^{j\frac{2\pi k}{N}}) = X_n(k) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-j\frac{2\pi km}{N}} \\ (0 \leq k \leq N-1).$$

与频域分析相关的演示界面如图 3 所示.

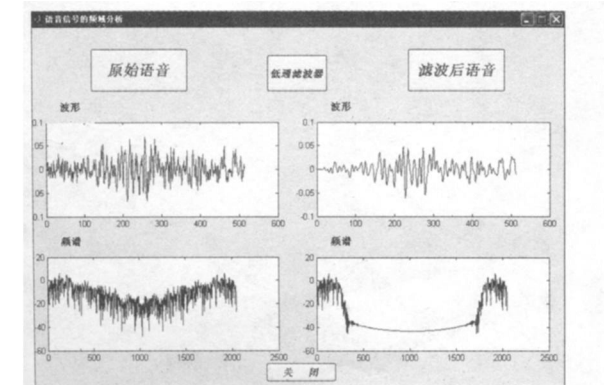


图 3 语音信号的频域分析

该演示平台取其中一帧语音信号,对其取离散的短时傅立叶变换,可以使学生直观地看到语音信号离散的短时傅立叶变换后得到的频谱形状,以加深对相关概念的理解.同时,该演示程序还生成了该语音信号经过一个低通滤波器以后的波形,可以形象地展示出语音信号经过低通滤波器前后波形和频谱的变化情况.

4 语音信号的线性预测分析

线性预测作为一种工具,几乎普遍地应用于语音信号处理的各个方面,是最有效和最流行的语音分析技术之一.线性预测分析的基本概念是:一个语音的抽样能够用过去若干个语音抽样的线性组合来逼近.通过使实际语音抽样与线性预测抽样之间差值的平方和达到最小值,就能够决定唯一的一组预测系数.

线性预测分析的基本原理是将被分析的信号用一个模型来表示,即将信号看做是某一个模型(即系统)的输出.这样就可以用模型参数来描述信号,如下所示.

$$u(n) \xrightarrow{H(z)} s(n)$$

根据 $H(z)$ 的不同,有三种不同的信号模型.其中最常用的是全极点模型,或称为自回归模型.可以推导出线性预测方程为:

$$\sum_{i=1}^p a_i \phi(j-i) = \phi(j-0) \quad (j=1, 2, \dots, p),$$

其中, $\phi(j-i) = \sum_n s(n-j)s(n-i)$. 而最小均方误差为 $E = \phi(0,0) - \sum_{i=1}^p a_i \phi(0,i)$. 具体推导过程省略.

而线性预测分析的关键就是求解线性预测方程,得到线性预测系数 $a_i (i=1, 2, \dots, p)$. 这里采用的解法是自相关法,可以把线性预测方程改写为:

$$\sum_{i=1}^p a_i r(|j-i|) = r(j) \quad (1 \leq j \leq p),$$

其中, $r(j)$ 表示语音信号的自相关函数.而最小均方误差也变为 $E = r(0) - \sum_{i=1}^p a_i r(i)$. 对于改写后的线性预测方程,可以采用递归方法求解,最常用的递归求解方法是莱文逊-杜宾(Levinson-Durbin)算法.

这个算法的过程和步骤为:

- (1) 对于 $i=0$ 时, $E_0 = r(0)$;
- (2) 对于第 i 次递归:

$$k_i = \frac{1}{E_{i-1}} \sum_{j=0}^{i-1} a_j^{i-1} r(j-i) \quad (1 \leq i \leq p),$$

$$a_i^i = k_i,$$

$$a_j^i = a_j^{i-1} - k_i a_{i-j}^{i-1} \quad (1 \leq j \leq i-1),$$

$$E_i = (1 - k_i^2) E_{i-1};$$

(3) 最终 $a_j = a_j^p, 1 \leq j \leq p$

语音信号的线性预测分析的演示界面如图 4 所示.

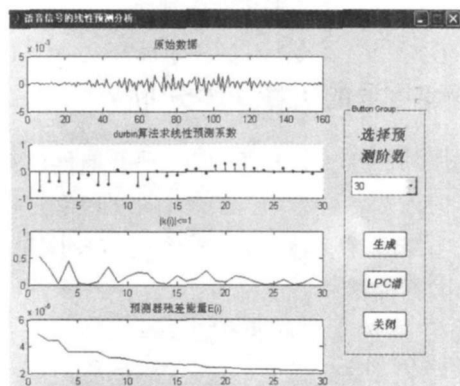


图 4 语音信号的线性预测分析

5 结语

本文围绕着语音信号的几种分析方法,介绍了

一个利用 MATLAB 开发的语音信号处理课程教学演示平台^[4]. 利用此教学演示平台,可以非常清楚地看到各种演示波形,并可通过改变参数观察其波形变换过程. 在实际的教学过程中可以看到,采用此教学演示平台有助于学生形象化地理解教学内容,极大地调动了学生的学习积极性,取得了比较好的教学效果.

参 考 文 献

- [1] 胡航. 语音信号处理 [M]. 哈尔滨:哈尔滨工业大学出版社, 2003
- [2] 张力. MATLAB 在语音信号处理辅助教学中的应用 [J]. 电气电子教学学报, 2005 (2): 96 - 99.
- [3] 蔡莲红, 黄德智. 现代语音技术基础与应用 [M]. 北京:清华大学出版社, 2003
- [4] 刘庆华, 陈紫强. 基于 MATLAB 和 DSP 的语音信号处理课程的建设 [J]. 电气电子教学学报, 2006 (4): 26 - 28

Design of the voice signal processing teaching platform based on MATLAB

WANG Zhi-peng, ZHANG Lei

(School of Physics and Electronic Engineering, Nanyang Normal University, Nanyang 473061, China)

Abstract: Based on the main features of the course of Speech Signal Processing, this paper adopts MATLAB to introduce a teaching demonstration platform. This platform will enable the students to understand and master the basic principles and analytical methods of this course.

Key words: the speech signal processing; MATLAB; time domain analysis; frequency domain analysis; linear prediction analysis