ISSN 1009-3044
Computer Knowledge and Technology 电脑知识与技术
Vol.5,No.22, August 2009, pp.6280-6282

E-mail: eduf@cccc.net.cn
http://www.dnzs.net.cn
Tel:+86-551-5690963 5690964

语音信号的预处理和特征提取技术

张 节

(武汉科技大学 信息科学与工程学院,湖北 武汉 430081)

摘要:语音信号处理是一门多学科交叉的综合学科,它包含了语音学和数字信号处理等基础学科。这篇文章对语音信号作了两个方面的研究:语音信号的预处理和语音信号的特征提取。预处理是为了更好地进行语音信号的特征提取,是语音信号特征提取获得成功的重要保障。语音信号的预处理介绍了信号的主分量分析(PCA)技术和白化(whitening)技术.而语音信号的特征提取分为时域的和频域的特征提取。并用 Matlab 编程实现了一段语音信号的分析处理。

关键词:语音信号;预处理;PCA;白化;特征提取;时域;频域

中图分类号:TP391 **文献标识码**:A **文章编号**:1009-3044(2009)22-6280-03

The Technology of Speech Signal Pre-processing and Feature Extraction

ZHANG Jie

(Wuhan University of Science and Technology, Wuhan 430081, China)

Abstract: Speech signal processing is a multi-disciplinary subject intersected by phonetics, digital signal processing, and so on. The article did the research about speech signal pre-processing and feature extraction. Pre-processing is prepared for the extraction, and guaranteed for the success of the feature extraction. The first part introduced the PCA and Whitening of the speech signal. The other part discussed the speech signal's time-domain and frequency-domain's features, and used Matlab to analyze one section of speech signal.

Key words: speech signal; pre-processing; PCA; whitening; feature extraction; time domain; frequency domain

语音是由人类发音器官发出的且具有一定意义的声音。声音中包含有一定的意义是语音同其它声音的本质区别。因此研究语音信号中所包含的意义,也即语音信号的特征是语音信号研究的重要目标。同时语音信号也具有和其它信号共有的一些特征参数,比如周期,频率,能量等,也需要使用它们来分析。

1 语音信号的预处理

语音信号的预处理也叫作前端处理,是指在特征提取之前,先对原始语音进行处理,使处理后的信号更能满足实际的需要,对提高处理精确度有重要的意义。

1.1 主分量分析技术(Principal Component analysis)

主分量分析技术是设法将原来的多个指标化为少数的几个主要指标的统计分析方法, 并且少数的主要指标要尽量多的反映原来多个指标的信息,还要求它们是相互独立的。从数学的角度来看,这是一种降维的处理技术。而求主分量的方法中最简单的是取原来指标的线性组合,然后调整组合系数。例如:如果有 p 个指标(x_1,x_2,……x_p)将它们通过一定的方法线性组合后变为 m 个指标(z_1,z_2 ……z_m),且

$$\begin{cases} z_1 = l_{11}x_1 + l_{12}x_2 + \ldots + l_{1p}x_p \\ z_2 = l_{21}x_1 + l_{22}x_2 + \ldots + l_{2p}x_p \\ \cdots\cdots \\ z_m = l_{m1}x_1 + l_{m2}x_2 + \ldots + l_{mp}x_p \end{cases}, \quad m \le p$$

很明显变量个数减少了,简化了运算,抓住了主要的矛盾。

1.2 白化技术(Whitening)

白化处理可以减少待估计的参数个数,降低分析问题的难度。白化的主要原理是对观测信号 X 进行线性变换,Z=WX,使所得的新向量 Z 之间互不相关,即 Z 的协方差矩阵为单位阵。$Rz=E\{Z^*Z^T\}=I$,其中 W 为白化矩阵。在一些信号处理领域如盲信号分离中,先对信号白化处理,可以有效地减少计算量。

2 语音信号的特征提取

人耳能够听到的音频频率范围是 60Hz-20KHz,其中语音大约分布在 300Hz-4KHz 之内。语音信号是连续模拟信号,而计算机只能处理数字化信息,所以要提取语音信号的特征必须先将语音信号数字化后才能在计算机上进行处理.并且语音信号是一种非平稳随机信号,只有在很短的时间内才认为是变化缓慢的,在这个短的时间区间内音频特征保持稳定,此时提取的特征才是有意义的.

2.1 语音信号的时域特征提取

语音信号时域可以提取的短时信息有:短时平均能量,短时过零率和线性预测系数。首先对语音信号进行采样,得到 K 个采样点,再把得到的采样点分割成前后迭代的音频帧(相邻帧之间的迭加率一般为 30%-50%)。短时平均能量指在一个短时音频帧内采

收稿日期:2009-04-10

样点所聚集的能量,它可以方便地表示整个时间段内幅度的变化。其定义为:

$$STE(t)=\frac{1}{N}\sum_{n}^{N}\left|f_t(n)\right|^2$$

短时平均能量可以直接应用到有声/静音检测中.若某一短时帧的平均能量低于某一个事先设定的阈值,则为静音,否则为非静音。

过零率是指在一个短时帧内,离散采样信号值由负到正和由正到负的次数.它可以有效的刻画不同的音频信号。其定义为:

$$ZCR(t)=\frac{1}{2}\sum_{n}^{N}\left|sign(f_t(n))-sign(f_t(n-1))\right|$$

其中,

$$sign(f_t(n))=\begin{cases}1, x(n)\geq 0\\ -1, x(n)<0\end{cases}$$

对于语音信号,浊音信号的过零率低,清音信号的过零率高.语音信号的开始和结束都大量集中了清音信号,所以在语音信号中,开始和结束的过零率会明显地升高,因此可以用过零率区分清音和浊音,以及语音信号的开始和结束。

线性预测的基本原理是把待分析的信号用一个模型表示出来,信号是这个模型的输出,构成这个模型的参数是这个信号的重要特征,称为线性预测系数。参数的确定是根据均方误差最小而得到的。

2.2 语音信号的频域特征提取

把原始信号先进行傅立叶变换转换到频域,再分析信号的特征,称为频域特征提取。频域特征主要有:LPC 倒谱系数和 Mel 系数。

LPC 倒谱系数的提取过程为:首先用数字滤波器对音频帧所包含的采样点进行预加重处理,对预加重处理后的音频帧内信号加窗函数,然后对它进行自相关分析,把这个结果施以 P 阶线性预测计算,得到长度为 P 的序列 X_p,得到音频帧的 LPC 倒谱系数。LPC 倒谱系数可以用来区分语音和非语音信号。

Mel 系数是建立在傅立叶和倒谱分析的基础上的。对短时音频帧上的采样点进行傅立叶变换,得到这个短时音频帧在每个频率上的能量。将整个频率分为 n 个就形成了 MFCC(Mel 系数)。如果对提取出来的 Mel 系数再计算其对应的倒谱系数,就是 Mel 倒谱系数。

3 语音信号特征提取实验

用软件录取一段 wav 格式的语音文件 so.wav,然后用如下代码(计算短时平均能量和短时过零率)在 Matlab 上运行(将 5 个程序文件与语音文件放入同一个文件夹内运行):

第一个文件(定义了 En 函数):

```
function En = energy(x,wintype,winamp,winlen)
error(nargchk(1,4,nargin,´struct´));
win = (winamp*(window(str2func(wintype),winlen))).´;
x2 = x.^2;
En = winconv(x2,wintype,win,winlen);
```

第二个文件(定义了 sgn 函数):

```
function y = sgn(x)
y = (x>=0) + (-1)*(x<0);
```

第三个文件(定义了 winconv 函数):

```
function y = winconv(x,varargin)
error(nargchk(1,4,nargin,´struct´));
len = length(varargin);
switch len
    case 0
        wintype = ´rectwin´;
        A = 1;
        L = length(x);
    case 1
        if ischar(varargin{1})
            wintype = lower(varargin{1});
            A = 1;
            L = length(x);
        end
    case 2
        if ischar(varargin{1}) && isreal(varargin{2})
            wintype = lower(varargin{1});
            A = varargin{2};
            L = length(x);
        end
    case 3
        if ischar(varargin{1}) && isreal(varargin{2}) &&...
                isreal(varargin{3})
```

```
            wintype = lower(varargin{1});
            A = varargin{2};
            L = varargin{3};
        end
End
w1 = (window(str2func(wintype),L)).´; A = A(:).´;
w = A.*w1;
NFFT = 2^(nextpow2(length(x)+L));
X = fft(x,NFFT); W = fft(w,NFFT);
Y = X.*W;
y = ifft(Y,NFFT);
```

第四个文件(定义了 zc 函数):

```
function zc = zerocross(x,wintype,winamp,winlen)
error(nargchk(1,4,nargin,´struct´));
x1 = x;
x2 = [0, x(1:end-1)];
firstDiff = sgn(x1)-sgn(x2);
absFirstDiff = abs(firstDiff);
zc = winconv(absFirstDiff,wintype,winamp,winlen);
```

第五个文件(主要文件):

```
[x,Fs] = wavread(´so.wav´);
x = x.´;
N = length(x);n = 0:N-1;
ts = n*(1/Fs);
wintype = ´rectwin´;
winlen = 201;
winamp = [0.5,1]*(1/winlen);
zc = zerocross(x,wintype,winamp(1),winlen);
E = energy(x,wintype,winamp(2),winlen);
out = (winlen-1)/2:(N+winlen-1)-(winlen-1)/2;
t = (out-(winlen-1)/2)*(1/Fs);
figure;
plot(ts,x); hold on;
plot(t,zc(out),´r´,´Linewidth´,2); xlabel(´t, seconds´);
title(´Short-time Zero Crossing Rate´);
legend(´signal´,´STZCR´);
figure;
plot(ts,x); hold on;
plot(t,E(out),´r´,´Linewidth´,2); xlabel(´t, seconds´);
title(´Short-time Energy´);
legend(´signal´,´STE´);
```

实验结果如图 1、图 2。

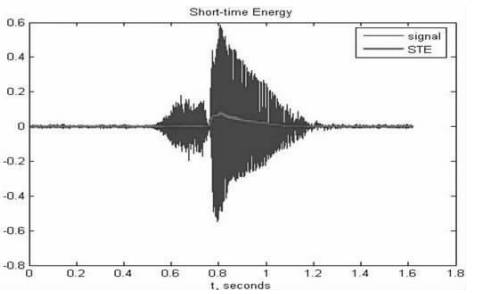

图 1

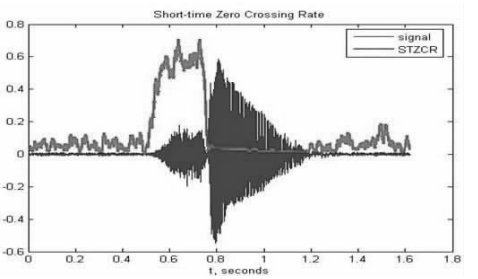

图 2

4 结果分析

从上面的第一幅图可以看出语音信号的能量在 0.85s 左右达到峰值,然后逐渐下降,1.2s 左右恢复为静音,清音(前一部分)的短时能量低,浊音(后一部分)的短时能量高.从第二幅图中可以看出 0.5s 过零率升高并达到峰值,0.8s 时下降为 0,清音(前一部分)的短时过零率高,浊音(后一部分)的短时过零率低。(语音信号为英文单词 so 的发音)

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,2002.
[2] 韩纪庆.语音信号处理[M].北京:清华大学出版社,2004.
[3] 程沛青.数字信号处理[M].北京:清华大学出版社,2007.
[4] 马建仓,牛奕龙,陈海洋.盲信号处理[M].北京:国防工业出版社,2006.
[5] 张发启.盲信号处理与应用[M].西安:西安电子科技大学,2006.
[6] 葛哲学.精通 MATLAB[M].北京:电子工业出版社,2008.
[7] 张志涌.精通 MATLAB6.5 版[M].北京:北京航空航天出版社,2003.
[8] 徐济仁,陈家松,徐屹.语音信号预处理技术综述[J].计算机应用,2001(27):26-28.
[9] 刘静萍,姜占财,德西嘉措.语音信号的预处理技术探讨[J].甘肃联合大学学报,2006(20):61-64.

张节(1985-),男,湖北武汉人,硕士在读,主要研究方向为语音信号,盲信号处理等。