

# 用LPC及DTW进行语音模式比较的设计与实现

范新伟<sup>1</sup>, 申瑞民<sup>1</sup>, 杜彦蕊<sup>2</sup>

(1. 上海交通大学计算机科学与工程系, 上海 200030; 2. 哈尔滨工程大学信息与通信工程学院, 哈尔滨 150001)

**摘 要:** 阐述了提取LPC特征参数和DTW模式匹配基本原理, 并进一步实现了求解LPC的杜宾算法和有效进行DTW的一个算法。在此基础上, 根据英语学习的实际需要, 设计并实现了一个语音模式比较实例。

**关键词:** 线性预测编码; 动态时间规整; 模式匹配; 语音识别

## Design and Implementation of Speech Pattern Comparison by Using LPC & DTW Method

FAN Xinwei<sup>1</sup>, SHEN Ruimin<sup>1</sup>, DU Yanrui<sup>2</sup>

(1. Department of Computer Science and Engineering, Shanghai Jiaotong Univ., Shanghai 200030;

2. Information and Communication Engineering College, Harbin Engineering University, Harbin 150001)

**【Abstract】** This paper expatiates the basic principles of the LPC characteristic coefficients and the DTW pattern matching, then implements the Durbin algorithm for the LPC and an efficient algorithm for the DTW. Basing on the requirement of English study, it designs and implements a speech pattern comparison application.

**【Key words】** Linenar prediction coding; Dynamic time warping; Pattern matching; Speech recognition

### 1 LPC, DTW的原理及实现算法

#### 1.1 LPC线性预测编码

LPC (linear prediction coding) 线性预测编码是一种重要的语音分析技术, 也是一种非常有效的低比特率高质量语音的编码方法。LPC有一个假定: 即声音信号是一个按时间顺序的线性序列, 那么在时间n时刻, 信号S(n)应该能够用一个过去的p个时刻的信号组成的线性组合来逼近。

$$S(n) \approx a_1 S(n-1) + a_2 S(n-2) + \dots + a_p S(n-p) \quad (1)$$

取S(n)的估计值  $\hat{S}_n = \sum_{i=1}^p a_i S(n-i)$ , 估计误差为

$$e(n) = S(n) - \hat{S}_n = S(n) - \sum_{i=1}^p a_i S(n-i) \quad (2)$$

现在定义  $E_n = \sum_{m=1}^n e_m^2$ , LPC基本的问题就是寻找一组参数使得短时误差平方和最小。

令  $S_n(m) = S(m+n)$ ,  $e_n(m) = d(m+n)$

$$E_n = \sum_m (S_n(m) - \sum_{i=1}^p a_i S_n(m-i))^2 \quad (3)$$

令:  $\frac{\partial E_n}{\partial a_i} = 0$ ,  $i = 1, 2, \dots, p$  得到

$$\sum_m (S_n(m) - \sum_{i=1}^p a_i S_n(m-i))^2 = \sum_{k=1}^p \hat{a}_k$$

$$\sum_m S_n(m-i) S_n(m-k) \quad 1 \leq i \leq p \quad (4)$$

其中  $\hat{a}_k$  是使  $E_n$  最小的  $a_k$  的值。再令:

$$\Phi_n(i, k) = \sum_m S_n(m-i) S_n(m-k)$$

于是式(4)改为

$$\sum_{k=1}^p a_n \Phi_n(i, k) = \Phi_n(i, 0) \quad (5)$$

最后当  $E_n$  取最小值时可以表示为

$$E_n = \Phi_n(0, 0) - \sum_{k=1}^p a_k \Phi_k(0, k) \quad (6)$$

只要能够解出式(6), 就能够得出使估计误差平方和最小的一组参数  $\hat{a}_k$ ,  $k = 1, 2, \dots, p$ 。这一组参数就是LPC线性预测编码的特征参数。

#### 1.2 LPC参数求解——杜宾算法

LPC特征参数主要的求解方法有自相关法 (autocorrelation method)、协方差法 (covariance method) 这里主要介绍自相关法中的杜宾算法 (以10阶为例)。

杜宾算法求解矩阵为

$$\sum_{i=1}^{10} a_i r'(|i-k|) = -r'(k) \quad k=1, \dots, 10$$

其中  $r'$  是经过调整的自相关系数, 其中的调整方法是针对语音识别的具体问题经过加窗函数等调整以提高效率, 实现细节在后面的系统实现过程中详细阐述。杜宾算法具体为:

$$\begin{aligned} E^{[0]} &= r'(0) \\ \text{for } i &= 1 \quad \text{to} \quad 10 \\ a_0^{[i-1]} &= 1 \\ k_i &= - \left[ \sum_{j=0}^{i-1} a_j^{[i-1]} r'(i-j) \right] / E^{[i-1]} \\ a_i^{[i]} &= k_i \\ \text{for } j &= 1 \quad \text{to} \quad i-1 \\ a_j^{[i]} &= a_j^{[i-1]} + k_i a_{i-j}^{[i-1]} \\ \text{end} \\ E^{[i]} &= (1 - k_i^2) E^{[i-1]} \\ \text{end} \end{aligned}$$

**作者简介:** 范新伟 (1978 - ), 男, 硕士生, 主要研究网络教育标准、语音识别、多媒体; 申瑞民, 博导; 杜彦蕊, 硕士生

收稿日期: 2003-01-20

E-mail: xwfan@sjtu.edu.cn

最后LPC参数为

$$a_j = a_j^{[10]}, j = 0, \dots, 10, a_0 = 1.0$$

### 1.3 DTW动态时间归整匹配

DTW (dynamic time warping) 动态时间规整匹配, 基于动态规划的思想, 解决了发音长短不一的匹配问题, 是语音识别中出现较早、较为经典的一种算法。

语音信号在对每一帧提取LPC特征参数之后, 就转化成了一组LPC特征向量。语音识别就是要将这个特征向量同模式库中已存的语音特征向量(参考模式)进行模式匹配, 寻找距离最短的模式。由于读入的语音信号和参考模式语音长短不同, 因此需要通过DTW动态计算两个长度不同的模式之间的相似程度, 或者叫做失真距离。

假设待测模式总共有n帧, 分别为1, 2, ..., n帧; 参考模式共有N帧, 分别为1, 2, ..., N帧。将待测模式各帧沿二维坐标的横轴排列, 参考模式各帧沿二维坐标的纵轴排列。将坐标中横轴每一帧与纵轴的每一帧横纵相连形成一个网格。每一格表示待测模式的某一帧和参考模式的某一帧之间的距离, 即两个向量之间的欧氏距离

$$d(x, y) = \sqrt{\sum (x_i^2 - y_i^2)} \quad (7)$$

DTW就是在这些网格中找到一条累加距离最短的路径, 即最短的欧氏距离之和, 其结果就是DTW的结果。当然这个匹配路径需要服从一定的约束: 匹配路径不能回退; 待测模式和参考模式的每一帧都必须匹配到; 路径中每个方格内的欧氏距离都要累加到全局距离(距离和)中。

下面介绍一种高效的算法来求解(假设行号和列号都是从0开始计数的), 如图1。

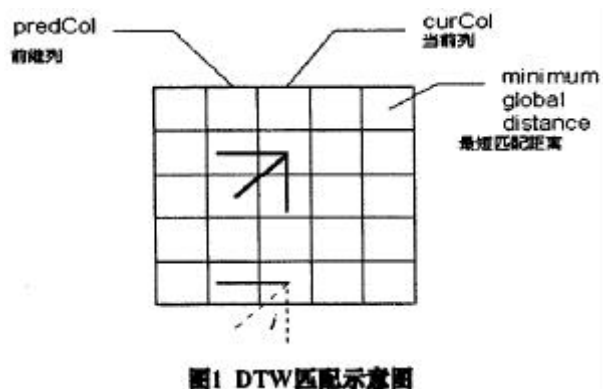


图1 DTW匹配示意图

方格(i, j)在路径中可能的前继方格可能有3个; 而方格(i, 0)只能有一个。

(1) 从第0列的最下面一个方格开始计算, 那么全局匹配距离在这里就是这个方格的欧氏距离。该列其他的方格的全局匹配距离就等于下面一个方格的全局匹配距离加上该方格的欧氏距离。这一列作为前继列(predCol)。

(2) 计算连续的下一列的j全局匹配距离, 仍然从最下面一个方格开始计算。最下端的方格的全局匹配距离就等于前继列的最下端的方格的全局匹配距离加上该方格的欧氏距离。这一列称为当前列(curCol)。

(3) 计算当前列的其余的方格的全局匹配距离。例如方格(i, j)的全局匹配距离就是它的欧氏距离加上方格(i-1, j-1), (i-1, j), (i, j-1)中最小的全局匹配距离。

(4) 将当前列(curCol)改名为前继列(predCol), 并且返回(2)递归计算。

最后计算出最后一列最上端的方格的全局匹配距离就是DTW所计算出的最小匹配距离也就是需要的失真距离。

## 2 语音模式比较系统设计

### 2.1 系统需求

语音模式比较在英语学习中应用的目的不同于其他典型的语音识别。语音识别的目的是在模式库中所有的模式中进行匹配, 找到和待测试的模式最为接近最为匹配的那个模式, 从而判断出测试者所讲的语音的含义。而在英语学习中, 语音模式比较的目的是将测试者的发音和模式库中指定的某个具体的参考模式进行比较, 判断其发音是否符合其标准发音, 甚至给出改进发音的建议。

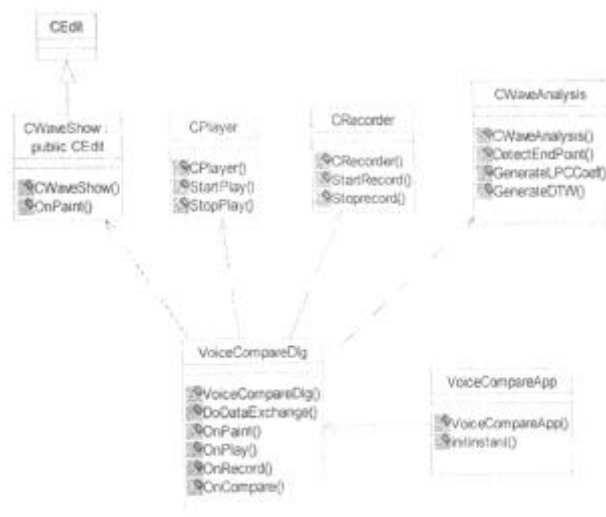


图2 设计类图

### 2.2 设计类图

本系统采用VC++开发, 设计类图及主要方法如图2。

VoiceCompareDlg类是进行语音模式比较的主界面, 它的成员包括CwaveShow、Cplayer、Crecorder、CwaveAnalysis各类的实例。分别用于显示波形、语音播放、语音录入、语音匹配的功能。CWaveAnalysis该类是语音模式比较的核心类, 它根据录入的语音进行分帧、端点检测, LPC特征向量提取, 以及给出最后的DTW匹配结果。

## 3 语音模式比较系统的实现

### 3.1 系统常量描述

语音采样率为PCM 8 000 kHz, 16位, 单声道, 25 kB/m。取10 ms为一帧, 假设10 ms内语音的特征保持不变。每次步长为5 ms即半帧, 以此步长前进, 降低误差。

### 3.2 语音录入

语音录入通过标准麦克风读入, 按上述常量描述的采样率和格式进行。

### 3.3 波形整理

将数字信号按照短整型整数读入, 整理数据中的异常点, 将异常的数据削减成常规值。

### 3.4 端点检测

采用短时域分析方法, 通过短时能量以及短时过零率来判断端点, 区分发音区和静音区。短时能量需要加窗函数, 本例采用海明窗(定义域外为零):

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)W(n-m)]^2 = \sum_{m=-\infty}^{\infty} x^2(m)h(n-m) \quad (8)$$

$$h(n)=0.54-0.46\cos\frac{2\pi n}{N-1}) \quad 0 \leq n \leq N-1 \quad (9)$$

计算每一帧的短时能量，可以通过取经验值得到静音部分的能量范围。判断每一帧的短时能量是否在静音能量范围之内从而得出这一帧是否是有声帧，只有连续几帧都是有声帧的情况才认为是找到了有声部分的起始结束端点。再通过过零率的细微调整最后确定端点。

过零率就是数字信号正负相接往返的次数。定义：

$$Z_n = \sum_{m=-\infty}^{\infty} |Sgn[x(m)] - Sgn[x(m-1)]| W(n-m) \quad (10)$$

$$Sgn[x(n)] = \begin{cases} 1 & x(n) \geq 0 \\ -1 & x(n) < 0 \end{cases} \quad (11)$$

$$W(n) = \begin{cases} \frac{1}{2N} & 0 \leq n \leq N-1 \\ 0 & \text{other} \end{cases} \quad (12)$$

用过零率调整端点的依据就是有声部分过零率比较低，无声部分较高。通过对短时能量找到的端点先后几帧过量率的检验最后确定端点所在的位置。

### 3.5 LPC参数向量的提取——杜宾算法的具体实现

针对每一帧提取10阶LPC参数向量，需要用到待提取帧先后共3帧240个数据进行自相关运算。于是定义加窗函数，LP分析窗包括两部分：第一部分是半个海明窗；第二部分是余弦函数。如下：

$$wlp(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{399}\right) & n = 0, \dots, 199 \\ \cos\left(\frac{2\pi(n-200)}{159}\right) & n = 200, \dots, 239 \end{cases} \quad (13)$$

这240个共3帧的数据中，40个（半帧）来自于当前帧的后续，120个（一个半帧）来自于当前帧的前续。于是加窗后的信号：

$$s' = wlp(n)s(n) \quad n = 0, \dots, 239 \quad (14)$$

$$\text{自相关系数: } r(k) = \sum_{n=k}^{239} s'(n)s'(n-k) \quad k = 0, \dots, 10 \quad (15)$$

为了避免输入信号过低产生数学问题， $r(0)$ 必须有一个下限 $r(0)=1.0$ 。于是自相关系数调整为：

$$\begin{aligned} r'(0) &= 1.0001 \quad r(0) \\ r'(k) &= wlag(k)r(k) \quad k = 1, \dots, 10 \\ wlag(k) &= \exp\left[-\frac{1}{2}\left(\frac{2\pi f_0 k}{f_s}\right)^2\right] \quad k = 1, \dots, 10 \end{aligned} \quad (16)$$

对调整后的自相关系数应用杜宾算法之后，就得出该帧的LPC特征参数向量为 $(a_1, a_2, \dots, a_{10})$ 。

### 3.6 应用DTW计算匹配距离

提取出了每一个有声帧的LPC特征向量，就可以和模式

（上接第125页）

机结合，创建面向非专业人士的服装CAD系统是一个很有创意的想法。款式的自动设计目前国内外仍然没有相关的成熟系统，具有一定的创新性和研究价值。由于设计本身的特点决定了其知识表达的困难及设计系统构成技术路线上的多种可能性。本文给出了利用结合专业知识的GA寻优，并采用RBF人工神经网络逼近GA的适合度函数。提供了一条有效地解决款式自动设计问题的途径。

### 参考文献

1 Miller P B. AutoCAD for the Apparel Industry. Delmar Publishers Inc. Albany, N Y, 1994

库中指定的参考模式（LPC特征向量）相比较，得出测试者的发音和标准发音的距离。应用前面提到的DTW的的算法

可以递归地求解出最后的模式匹配的距离（失真距离）。

### 3.7 结果反馈

根据DTW计算出的最小匹配距离对照经验值得出测试者得发音是否符合标准发音，以及符合的程度。给出A,B,C,D这4个等级。图3就是本例在得出结果时的反馈。可以看到，上端的标准语音波形和下端的测试者录入的语音波形形状上相似并且匹配的距离（失真距离）很小。所以在左下的打分给出了A的等级。

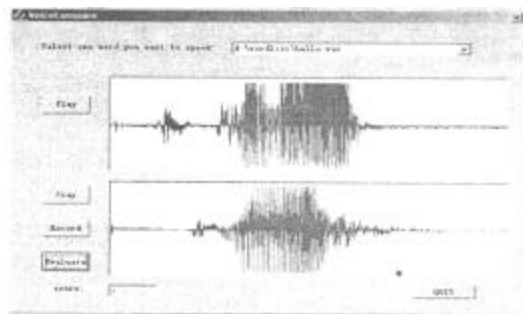


图3 运行界面

## 4 结束与展望

至此，一个通过提取LPC特征参数和DTW进行动态时间规整模式匹配的语音模式比较系统就完成了。还有许多问题需要继续改进，例如匹配的结果的精确度还可以进一步提高，参数提取的阶数可以进一步增大等。

设想：经过大量的经验参数的积累，比如大量的学习者的发音特征，结合数据挖掘的聚类算法得出不同发音特点的聚类，再经过专家对各类的发音评判，给出各聚类的改进建议。这样在本系统给出最后的评分以后，可以进一步指出其发音的缺点并提出改进建议，这样的效果会更好。

### 参考文献

- 1 ITU-T G.729. Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction(CS-ACELP). <http://www.laogu.com/download/g729.pdf>
- 2 Rabiner L, Huang Juangbiing. Fundamentals of Speech Recognition. Pearson Education POD, 1993
- 3 Wrigley S N. Speech Recognition by Dnamic Time Warping. <http://www.dcs.shef.ac.uk/~stu/com326/index.html>
- 4 谢依兰. 语音讯号位数处理. 台湾: 松岗电脑图书资料股份有限公司, 1992
- 5 胡光锐. 语音处理与识别. 上海: 上海科学技术文献出版社, 1994

- 2 Ku I S.Com-puter-Aided Fashion Design.Kuyomunsa(in Korean), Seoul, Korea,1994
- 3 Nakanishi Y. Applying Evolutionary Systems to Design Aid System. Proceedings of Artificial Life V(Poster Presentation), 1996:147-154
- 4 Lee J Y, Cho S B. Interactive Genetic Algorithm for Content-based Image Retrieval. Proceedings of Asia Fuzzy Systems Symposium, 1998: 479-484
- 5 Kim H S, Cho S B. Application of Interactive Genetic Aalgorithm to Fashion Design. Engineering Application of Artificial Intelligence, 2000,13: 635-644