

语音倒谱特征的研究

王让定^{1,2}, 柴佩琪¹

(1. 同济大学人工智能研究室, 上海 200092; 2. 宁波大学信息科学与工程学院, 宁波 315211)

摘 要: 语音倒谱特征是语音识别中最常用的特征参数, 它表征了人类的听觉特征。该文在研究基于线性预测倒谱和非线性MEL刻度倒谱特征的基础上, 研究了LPCC和MFCC参数提取的算法原理及提取算法, 提出了一级、二级差分倒谱特征参数的提取算法。识别实验验证了MFCC参数的鲁棒性优于LPCC参数。

关键词: LPCC; MFCC; 特征提取

Research on Speech Cepstral Features

WANG Rangding^{1,2}, CHAI Peiqi¹

(1. AI Laboratory, Tongji University, Shanghai 200092; 2. College of Information Science and Engineering, Ningbo University, Ningbo 315211)

【Abstract】 The speech cepstral features are very important parameter in ASR, which symbolizes the property of HAS(human auditory system). The paper discusses cepstral features based on linear predictive and non-linear MEL scale, and presents LPCC and MFCC's extraction algorithm and differential coefficient and accelerated coefficient are also presented. The recognition experiment results show that MFCC's robustness is prior to LPCCs.

【Key words】 LPCC; MFCC; Feature extraction

当语音信号按Niquist采样频率采样后, 得到了数字化(一般按8位或16位量化)表示的语音。由于采样率一般在8kHz以上, 经数字化后的一段语音信号, 其数据量很大(尤其是在处理连续语音时), 不便于有效地、实时地分析处理语音。如何用能反映人类声道和听觉特性等的少量语音特征表示语音信息, 是语音编码、语音合成、语音识别、说话人确认和验证等技术中一直追求的目标。

几十年来, 人们研究了很多表示语音的特征参数, 如临界带特征矢量、线谱对(LSP)、线性预测系数(LPC)等。目前, 语音识别技术中应用最流行的特征参数是基于人的声道模型和听觉机理的LPCC(Linear Predictive Cepstral Coefficients)和MFCC(Mel Frequency Cepstral Coefficients)参数, 而MFCC参数在低频段具有较高的谱分辨率, 对噪声的鲁棒性优于LPCC, 因此更适合于语音识别。

本文研究了LPCC和MFCC参数提取的算法原理及提取算法, 提出了一级、二级差分系数 LPCC、LPCC和MFCC、MFCC的改进算法。通过实验, 研究分析了两类参数的鲁棒性。

1 LPCC参数的提取

在线性预测(LPC)分析中, 声道模型是式(1)表示的全极点模型^[1]

$$H(z) = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} = \frac{G}{A(z)} \quad (1)$$

式中p是LPC分析的阶数, a_j 为线性预测系数($j=1, 2, \dots, p$), $A(z)$ 称为逆滤波器。由式(1)可知LPC的分析就是求解P阶线性预测系数 a_j (本文通过自相关Durbin法求解)。假设求得的参数为 $(a_{ij})_{l \times p}$, 自相关矩阵为 $R(i, j)$, 其中 l 为分析的语音帧数, $1 \leq i \leq l$ 。

对式(1)两边取对数, 得

$$\hat{H}(z) = \ln G - \ln(1 - \sum_{l=1}^p a_l z^{-l}) = \sum_{k=-\infty}^{\infty} \hat{h}(k) z^{-k} \quad (2)$$

再对式(2)两边关于 z 求导, 有

$$\frac{-\sum_{n=1}^p n a_n z^{-n-1}}{1 - \sum_{l=1}^p a_l z^{-l}} = -\sum_{k=-\infty}^{\infty} k \hat{h}(k) z^{-k-1} \quad (3)$$

式(3)两边同乘 $-z(1 - \sum_{l=1}^p a_l z^{-l})$, 得

$$\sum_{n=1}^p n a_n z^{-n} = \sum_{n=-\infty}^{\infty} n \hat{h}(n) z^{-n} - \sum_{l=1}^p \sum_{k=-\infty}^{\infty} k \hat{h}(k) a_l z^{-k-l} \quad (4)$$

用 $l=n-k$ 代入式(4), 有

$$\begin{cases} n a_n = n \hat{h}(n) - \sum_{k=1}^{n-1} k \hat{h}(k) a_{n-k}, & 0 < n \leq p \\ 0 = n \hat{h}(n) - \sum_{k=n-p}^{n-1} k \hat{h}(k) a_{n-k}, & n > p \end{cases} \quad (5)$$

综上可得式(6)的LPCC倒谱的计算公式。

$$\hat{h}(n) = \begin{cases} 0 & n < 0 \\ \ln G & n = 0 \\ a_n + \sum_{k=1}^{n-1} \frac{k}{n} \hat{h}(k) a_{n-k} & 0 < n \leq p \\ \sum_{k=n-p}^{n-1} \frac{k}{n} \hat{h}(k) a_{n-k} & n > p \end{cases} \quad (6)$$

式中增益因子 $G = a_j R(j)$, $1 \leq j \leq p$ (这里仅考虑某一语音分析帧, 所以帧长下标未标注), 令 $C(n) = \hat{h}(n)$, 便得到了本文的倒谱特征的统一表示式。

2 MFCC参数的提取

语音的LPC参数在语音识别中具有最广泛的应用, 但它对噪声的影响特别敏感。人的耳朵能够从嘈杂的背景噪声中听到语音信号, 这是因为人的内耳基础膜对外来信号会产生调节作用。对不同的频率, 在相应的临界带宽内的信号会引

作者简介: 王让定(1962—), 男, 副教授、博士生, 主要研究方向: 语音识别, 语音编码, 计算机应用; 柴佩琪, 教授、博导
收稿日期: 2002-07-17

起基础膜上不同位置的振动。由此可用带通滤波器组来模仿人耳听觉，从而减少噪声对语音的影响。

(1) 算法原理^[2]

设语音信号的DFT为

$$X_a(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N} \quad 0 \leq k < N \quad (7)$$

其中 $x(n)$ 为输入的语音信号, N 表示傅立叶变换的点数。

我们定义一个 M 个滤波器组,采用的滤波器为三角滤波器,中心频率为 $f(m)$, $m=1,2,\dots,M$,则三角滤波器的频率响应按式(8)定义,频率响应波形如图1所示。

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{2(k-f(m-1))}{(f(m+1)-f(m-1))(f(m)-f(m-1))} & f(m-1) \leq k \leq f(m) \\ \frac{2(f(m+1)-k)}{(f(m+1)-f(m-1))(f(m+1)-f(m))} & f(m) \leq k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (8)$$

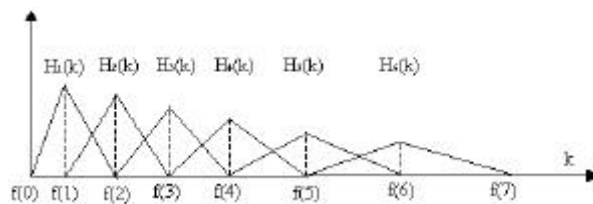


图1 用于计算Mel倒谱的三角滤波器

便于计算,本文将式(8)的三角滤波器简化为

$$H'_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) \leq k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (9)$$

其中 $\sum_{m=0}^{M-1} H'_m(k) = 1$ 。

Mel滤波器的中心频率定义为:

$$f(m) = \frac{N}{F_s} B^{-1} \left(B(f_h) + m \frac{B(f_h) - B(f_l)}{M+1} \right) \quad (10)$$

其中 f_h 和 f_l 分别为滤波器组的最高频率和最低频率, F_s 为采样频率,单位为Hz。 M 是滤波器组的数目, N 为FFT变换的点数,式中 $B^{-1}(b) = 700 \left(e^{\frac{b}{1125}} - 1 \right)$ 。

每个滤波器组的输出的对数能量为

$$S(m) = \ln \left(\sum_{k=0}^{N-1} |X_a(k)|^2 H'_m(k) \right) \quad , \quad 0 \leq m < M \quad (11)$$

经余弦变换得到MFCC系数

$$C(n) = \sum_{m=0}^{M-1} S(m) \cos(pn(m+0.5)/M) \quad , \quad 0 \leq n < M \quad (12)$$

(2) MFCC算法流程

设某语音信号为 $x(n)$,则算法处理流程为

- 1) 预加重 $x'_n = x_n - kx_{n-1}$, 其中 k 为预加重系数,一般取0.95;
- 2) 加窗(hamming窗), 帧长为 N ;
- 3) DFT变换;
- 4) 设计一个具有 M 个带通滤波器的滤波器组,采用三角滤波器,中心频率从 $0 \sim F_s/2$ 间按Mel频率分布;
- 5) 按式(11)计算每个滤波器组输出的对数能量;
- 6) 按式(12)求得MFCC系数。

3 差分特征参数的提取^[3]

在提取了LPCC和MFCC参数后,可统一用式(13)的差分特征参数提取算法提取 LPCC、LPCC和 MFCC、MFCC参数。

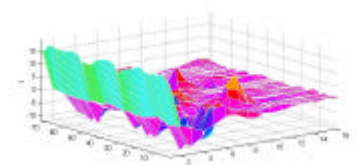
$$d_t = \begin{cases} c_{t+1} - c_t & t < \Theta \\ c_t - c_{t+1} & t \geq T - \Theta \\ \frac{\sum_{q=1}^{\Theta} q(c_{t+q} - c_{t-q})}{2 \sum_{q=1}^{\Theta} q^2} & \text{其它} \end{cases} \quad (13)$$

式中 d_t 表示第 t 个一阶差分倒谱系数, T 是为倒谱系数的维数,表示一阶导数的时间差,其值取1或2,1表示第 t 个倒谱系数。

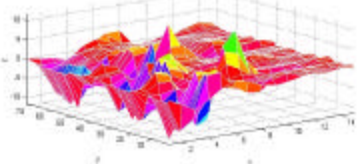
4 实验

4.1 实验一:语音特征参数提取

分析语音样本为汉语“左右为难”,采样率11025kHz,每个取样16bits,按本文讨论的算法提取的LPCC和MFCC参数如图2、图3所示,图中 x 轴表示倒谱系数的维数, y 轴表示语音分析的帧数, z 轴表示对应的倒谱值,LPCC参数阶数为16(倒谱系数为 $c_0 \sim c_{15}$),MFCC分析的滤波器组数为27,系数选取了前面的12个($c_0 \sim c_{11}$)。(a)图带有倒谱 c_0 项,(b)图不带倒谱 c_0 项,从图中可以看出,无论是LPCC系数还是MFCC系数,它们的第一维的值 c_0 的能量很大,故在一般识别系统中,将 c_0 称为能量系数,不作为倒谱系数中的一员。

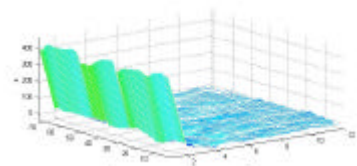


(a) 含倒谱 c_0 项

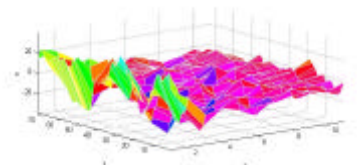


(b) 不含倒谱 c_0 项

图2 LPCC参数



(a) 含倒谱 c_0 项



(b) 不含倒谱 c_0 项

图3 MFCC参数

4.2 实验二：识别实验

本实验在两个识别平台上进行：一是我们自行开发的基于VQ的有限说话人有限汉语短语识别系统；二是在HTK3.1下训练的有限说话人有限汉语短语语音识别系统。实验的目的主要是验证所提语音参数的鲁棒性，由于我们的识别系统工作的背景环境中存在各种各样的干扰，输入系统的语音信号除了端点检测外，未进行语音增强处理，因此参数的鲁棒性直接影响识别精度。

实验中用于两个识别平台的训练和测试语音样本相同，共有27个汉语短语（2-4字），分别是开机、关机、正常、升温、降温、升x度，降x度（x表示1-9）等某设备控制命令。在带有噪声干扰的背景环境(SNR约为20dB)下采集，采集的语音样本来自3男2女，每个说话人对每个命令发音30次（分上午、下午和晚上采集，每个时间段各采集10次），共采集语音样本为 $5 \times 30 \times 27$ 个，其中70%用于训练，30%用于测试。两个识别平台分别采用LPCC+ LPCC+ LPCC和MFCC+ MFCC+ MFCC的33维特征参数，删去参数中的能量系数，每个识别参数为11维。在训练基于HTK的识别平台时，对每个汉语短语训练HMM(Hide Markov Model)模型，HMM状态为8个，混合分量数为3，流（stream）数依参数结构分设为1、2、3。两个识别系统分别以LPCC、MFCC、LPCC+ LPCC、MFCC+ MFCC、LPCC+ LPCC+ LPCC和MFCC+ MFCC+ MFCC结构形式的特征参数进行训练，产生的识别系统称为VQ_{LPCC}、VQ_{MFCC}、VQ_{LPCC+ LPCC}、VQ_{MFCC+ MFCC}和HTK_{LPCC}、HTK_{MFCC}、HTK_{LPCC+ LPCC}、HTK_{MFCC+ MFCC}，对所训练的不同系统用测试语音样本进行测试识别，同时对所训练的各识别系统进行了实时实验，其实验统计结果见表1所示。

表1中的实验结果（识别率）是按汉语短语识别的正确次数与总的待识短语总数之比统计的，表中实时1和实时2分别表示训练的和未训练的说话人的识别结果。从表中的结果可知，LPCC参数的鲁棒性比MFCC参数的鲁棒性差，尤其

（上接第30页）

号，并记录各线的信息；最后记录该盒片零件的总体信息，如该盒片零件的编号、名称等。各盒片零件的尺寸与其所在盒片的长、宽、高尺寸有着对应关系，因此其尺寸由所在盒片的长、宽、高来确定，对于尚未拼合的盒片零件，则先由系统给出基本尺寸，拼合后自动转换为其所属盒片的长、宽、高的对应值即可。

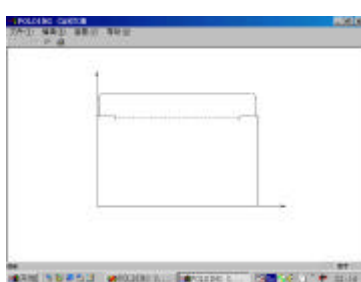


图3 盒片零件的表示

表1 实验结果

| 特征参数 | VQ (识别率) | | | | | | HTK (识别率) | | | | | |
|------|----------|------|------|------|------|------|-----------|------|------|------|------|------|
| | LPCC | MFCC | L+ | M+ | L+ + | M+ + | LPC C | MFCC | L+ | M+ | L+ + | M+ + |
| 测试 | 94.2 | 95.8 | 95.1 | 99.2 | 96.3 | 99.4 | 93.7 | 94.1 | 94.3 | 98.2 | 97.5 | 98.9 |
| 实时1 | 89.6 | 92.3 | 93.1 | 98.7 | 95.8 | 99.2 | 89.7 | 91.2 | 93.2 | 97.5 | 95.4 | 96.6 |
| 实时2 | 34.3 | 49.7 | 36.4 | 50.2 | 36.3 | 50.7 | 23.4 | 39.9 | 30.2 | 43.4 | 32.1 | 46.8 |

在带有背景噪声的环境下。另外，增加了一级和二级差分系数的识别参数，系统的总体性能得到了提高，但基于HMM模型的HTK语音识别系统的性能增加效果更为明显，其原因是采用多种流和混合分量训练的模型更具稳健性。

5 结论

本文系统地研究了反映语音倒谱特征的LPCC和MFCC参数的提取算法，提出了LPCC、MFCC一级、二级差分参数的提取方法。实验证明，文中研究的MFCC参数的提取算法提取的特征参数的稳健性好、识别率高。另外，通过实验也说明了MFCC参数的识别效率要比LPCC高。在本文研究的基础上，下一步重点研究MFCC各倒谱分量对识别率的影响，从而优化识别参数，以及基于感知（Perceptual）的LPCC和MFCC参数的特点和对识别系统的鲁棒性的影响。

参考文献

- 1 Rabiner L, Juang Biing-Hwang. Fundamentals of Speech Recognition. Prentice Hall,1993, 北京: 清华大学出版社(影印版),1999
- 2 Huang Xuedong, Acero A, Hon H W. Spoken Language Processing. Prentice Hall,2001
- 3 Young S, Kershaw D, Odell J, et al. The HTK Book.Microsoft Corporation & CUED,2000
- 4 Duda R O, Hart P E, Stork D G. Pattern Classification (Second Edition). A Wiley-interscience Publication, 2001
- 5 Wendt S, Fink G A, Kummert F. Forward Masking for Increased Robustness in Automatic Speech Recognition. In: Proc. of European Conf. on Speech Communication and Technology, Aalborg, Denmark, 2001,1:615-618
- 6 杨行峻. 语音数字信号处理. 北京: 电子工业出版社, 1995
- 7 Hermansky H. Perceptual Linear Predictive(PLP) Analysis for Speech. J Acoust Soc Am ,1990,87:1738-1752

4 结束语

基于拼合法的折叠纸结构CAD系统的结构设计功能已经开发完成，使用Folding Carton可以方便地设计和修改纸盒结构，设计的正确性也有一定的保证，提高了折叠纸盒结构设计的效率和质量。采用盒片零件拼成盒片的设计方法，可以单独处理成盒后的每一个盒面，为下一步对纸盒进行装潢设计创造了有利条件。

参考文献

- 1 王新成. 高级图形处理技术. 北京: 中国科学技术出版社, 2001
- 2 孔立钊. 计算机图形学系. 哈尔滨: 哈尔滨工业大学出版社, 2000
- 3 Sabin M. CAD System Component. Computer-aided Design, 1996, 28 (3): 183-192
- 4 Lu Liang. Folding Cartons with Fixtures: A Motion Planning Approach. IEEE Transactions on Robotics and Automation, 2000,16(8): 396